NOTES ON THE TEACHING OF STATISTICS IN SCHOOLS

By

B. C. Brookes, M.A.

WITH A FOREWORD BY

Professor E. S. Pearson
C.B.E., M.A., D.Sc.

ILLUSTRATED

HEINEMANN

BY THE SAME AUTHOR

An Introduction to Statistical Method

(HEINEMANN)

See back cover for description

NOTES ON THE TEACHING OF STATISTICS IN SCHOOLS

by

B. C. BROOKES, M.A.

Lately Senior Mathematics Master, Bedford Modern School
Member of the Teaching Committee of the
Royal Statistical Society

With a Foreword by

PROFESSOR E. S. PEARSON, C.B.E., M.A., D.Sc.

Department of Statistics, University College, London

WILLIAM HEINEMANN LTD
MELBOURNE : LONDON : TORONTO

FIRST PUBLISHED 1952

PUBLISHED BY
WILLIAM HEINEMANN LTD
99 GREAT RUSSELL STREET, LONDON, W.C.1
PRINTED IN GREAT BRITAIN BY BUTLER AND TANNER LTD.
FROME AND LONDON

Foreword

Figures are nowadays being used continually to prove this thing or that, yet faulty reasoning is all too common, even when the reasoner is trying to be honest. There can be no better way of encouraging a balanced adult view of figures than to make the boy or girl familiar at school with what has been termed the statistical approach. Developed by easy steps in which the children produce their own figures by observation, subject them to simple manipulation and speculate on their meaning, statistics can not only be made fascinating but also provide one of the best forms of elementary training in clear and critical thinking.

There are many teachers who are convinced that, given the opportunity, they could introduce in this way the simpler statistical ideas and problems to children of 14 or 15 as part of their general education and in so doing affect their intellectual approach to other subjects. Already, at a later stage, the occasional teaching of statistics as a more specialised Sixth Form subject, allied with mathematics, has been recognised by the inclusion of optional questions or papers in the General Certificate of Education syllabuses of several examining authorities. But as yet little help has been given to the pioneer teacher in the shape of a text-book or even an elementary set of Tables.

In their recent Report, the Royal Statistical Society's Committee on the Teaching of Statistics in Schools laid emphasis on these two aspects of the problem, the general and the specialised, and it was indeed in connection with discussions in this committee that Mr. Brookes first prepared the two sections of these 'Notes'. If Section II is a series of connected notes which should prove helpful in teaching for the G.C.E. syllabuses, Section I is really more than this. It seems to me to provide an admirable account of how one teacher with imagination would

approach those basic statistical concepts: variation, correlation, probability, sampling; how he would make his pupils collect the data by which to clothe these concepts with reality; and how he would use the results to encourage a critical and inquiring attitude of mind.

When the method of introducing a subject is still in the experimental stage, each teacher must to a great extent find the route best suited to the age and ability of his pupils, but there will be few who cannot make some use of the many and varied suggestions that Mr. Brookes has provided. Many of the ideas which he illustrates are not peculiar to statistics and he has realised the importance of taking every opportunity of linking up what is new with familiar ideas and methods already taught. Indeed, much will be gained if the statistical approach is seen to be largely a combination of common sense and scientific method under another name!

E. S. PEARSON.

Department of Statistics,
University College, London.
June 1952.

PREFACE

Reasons for teaching Statistics in schools have been given in a report prepared by the Teaching Committee of the Royal Statistical Society, which is obtainable from the Assistant Secretary of the Society, 4 Portugal Street, London, W.C.2. (Price, 1*s.*)

In these notes an attempt has been made to give practical detailed guidance to teachers in secondary schools on the teaching of Statistics at all levels. Section I describes work that might be included in the general curriculum: Section II is a teaching commentary on the published Statistics syllabuses of the G.C.E. examining authorities.

The notes are of course intended only for those who have had no academic training in, or practical experience of, the subject, but who nevertheless appreciate that Statistics is of both practical and educational value. The notes, moreover, are only suggestions: it is not pretended that they outline a complete or systematic course. But if by provoking constructive criticism they help to make Statistics more generally taught in schools, they will have served a useful purpose.

Professor E. S. Pearson kindly read Section I in manuscript and made helpful criticisms for which I am most grateful, but the responsibility for any deficiencies that remain is of course mine alone. I must also thank Mr. Cyril Bibby for his help and suggestions about the application of statistical methods to biology.

B. C. B.

CONTENTS

SECTION II

Teaching Notes on the G.C.E. Syllabuses in Statistics

SECTION I

The Teaching of Statistics in the General Curriculum

1. Introduction

The purpose of this Section is to give some teaching notes on the topics listed in Para. 22 of the Royal Statistical Society's *Memorandum on the Teaching of Statistics in Schools*. It is emphasised that these notes do not pretend to outline a systematic course of Statistics—for the greater part they merely suggest extensions of some of the topics that are already taught in the arithmetic course usually given to children of the age range 12 to 14 years in grammar, modern and technical schools. All that is proposed is that these statistical extensions can be discussed in class as suitable opportunities arise, as they frequently do, although an occasional review of a whole section, e.g. diagrams, is recommended as a means of consolidating previous work. It is also hoped that the notes will show convincingly that the suggested subject matter is not inherently difficult for schoolchildren and that it is valuable in relating the work of the mathematics classroom to other work and interests of children. In schools where the teacher of arithmetic is also responsible for the teaching of some science or geography, it will be easy for him to select examples for study that combine statistical interest with educational value. Elsewhere the mathematics teacher should try to enlist the co-operation of his scientific colleagues by persuading them to provide him with the results of class experiments; the numerical analysis can be carried out in the mathematics class and the discussion of the class results can follow in the science class—with benefit to both. The examples used in these notes are intended only to suggest to the teacher what can be done; it is important that he should use as far as possible data that directly concern the

members of his classes. Only if the class knows and understands the whole background from which the data have been collected can a discussion of the statistical implications be profitable. The emphasis of teaching Statistics in the general curriculum must be on interpretation rather than on technique.

2. Diagrams and Charts

The discussion of diagrams and charts could well begin by inviting a class to bring to school any interesting diagrams or charts that they notice in their casual reading of newspapers and popular periodicals. The material brought will almost certainly include a sufficient number of good and bad charts to provide the teacher with illustrative examples and to provoke the children to further search. The notes that follow may suggest to the teacher lines on which the subject may be discussed with 13- or 14-year-old children.

The purpose of a statistical diagram is to represent numerical data so that their salient features are more quickly and fully comprehended than by a study of the original numbers. If a diagram is to be helpful it must satisfy two conditions. First it must be simpler to comprehend than the numbers it represents, and secondly its picture of the numbers and their inter-relations must be a true one. Published diagrams sometimes fail by being over-elaborate, e.g. too complex or too highly coloured, or by being misleading, e.g. essentially incomplete or wrongly dimensioned.

Two main types of simple statistical tables will be considered: (*a*) Those showing a classification, e.g. the way in which the pupils of a school are divided into 'houses', (*b*) those giving a time-sequence of numbers, e.g. the daily attendance at a school. These we will call *class-data* and *time-data* respectively.

2(*a*) *Representation of Class-Data*

Class-data are commonly represented by one of three types of diagram known as the 'bar-chart', 'pie-chart' and 'symbol-chart'. The choice of chart depends partly on the aspect of the data to which attention is to be drawn. The three types will be

exemplified by means of the data of Table 1. The units selected for the bar-chart depend on the space available for the diagram. In Fig. 1 a centimetre represents the number 20. The figures for the pie-chart are obtained by multiplying each figure

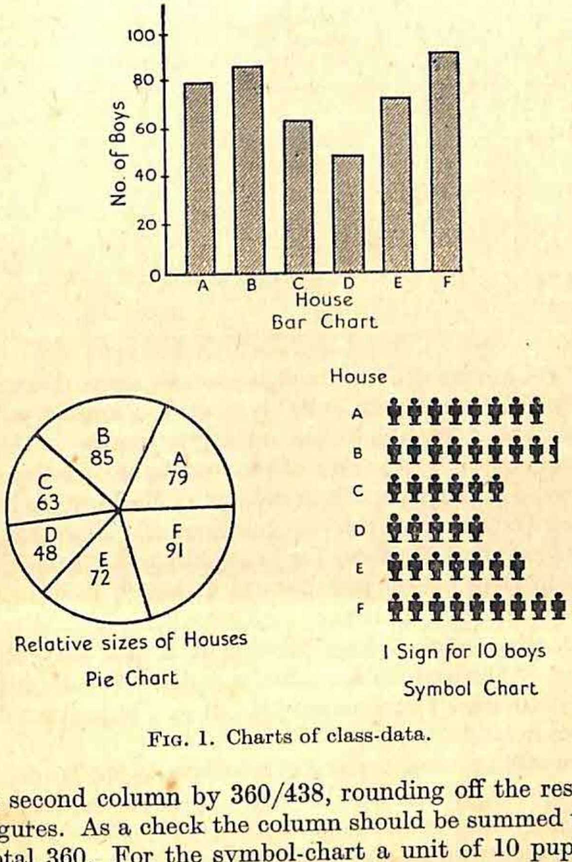

FIG. 1. Charts of class-data.

of the second column by 360/438, rounding off the results to two figures. As a check the column should be summed to give the total 360. For the symbol-chart a unit of 10 pupils has been chosen; fractions less than $\frac{1}{2}$ have been ignored, fractions greater than $\frac{1}{2}$ have been counted as units.

The bar-chart clearly shows the comparative sizes of the six

TABLE 1

Class-data

DISTRIBUTION OF PUPILS AMONG THE SIX HOUSES OF A SCHOOL

House	No. of pupils	Bar-chart (length of bar in cm.)	Pie-chart (angle of sector)	Symbol-chart (No. of units)
A	79	395	65°	8
B	85	425	70°	8½
C	63	315	52°	6
D	48	240	39°	5
E	72	360	59°	7
F	91	455	75°	9
Total	438		360°	

houses; the human eye is trained at an early age to discriminate between different lengths suitably placed. The sectors of the pie-chart are better able to show the relationship of the parts to the whole, i.e. of the size of any one house relative to the total school population; schoolchildren of 13–14 years have an unfailing judgment of the relative sizes of pie or cake. The 'Isotype' symbol-chart is, for unsophisticated minds, more attractive than the formal bar-chart though it is based on a similar principle. It is however difficult for a child to make an attractive symbol-chart because it is not easy to draw repeated symbols which are exactly similar. For display work a symbol at least 1 inch in length, cut as a stencil from card-board, is helpful.

There still occurs a type of symbol-chart that is misleading or ambiguous. When the different classes are represented by the same symbol in different sizes it is not usually possible to discover easily which dimension has been used as the basis of comparison. Areas are much more difficult than lengths to compare visually, and if the symbol is drawn to suggest something solid, e.g. a ship or a sack, there is even more difficulty

in making true comparisons. The checking of published examples of such symbol-charts provides useful exercises in ratios, square roots and cube roots for children who can use logarithms. Such exercises perform an educational function,

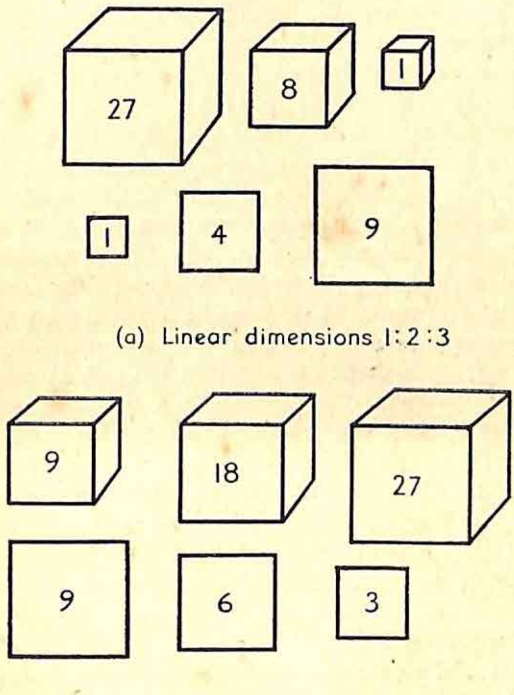

FIG. 2. Diagrammatic representations of the ratios 1 : 2 : 3.

if only they encourage children not to accept uncritically everything they see in print. Classroom models illustrating the ideas of Fig. 2 are essential to ensure that children appreciate the ambiguities of phrases like 'twice as big as' and 'three times the size of' applied to areas and volumes.

2(*b*) *Representation of Time-Data*

Time-data are represented by charts on which it has become conventional to mark 'time' along the base or x-axis and the other quantity along the y-axis. Examples of time-charts are commonly used to introduce the idea of co-ordinates before work is begun on graphs. (The word 'graph' is best applied to the diagram of a mathematical function.) It is therefore necessary only to mention here some important questions which are frequently overlooked.

First, should the points marked on the chart be joined by straight or smooth lines—or even joined at all? The answer depends partly on the nature of the quantity measured. If it is continuous, e.g. the temperature of a test-tube of melted wax that is being allowed to cool, then it is reasonable to draw a smooth curve through the plotted points. The fact that the curve is drawn implies that, at any instant during the period of the measurements, the wax has a temperature of which the ordinate corresponding to the instant is the best estimate. If the quantity measured is discrete, e.g. the daily attendance at the school, then the joining of successive plotted points is meaningless—though it is often done. If the quantity measured is continuous but erratic and is measured at intervals, e.g. the barometric height or the depth of a reservoir measured daily, then it is reasonable to join successive points by straight lines. The straight lines imply continuity of the variable but, on the part of the observer, ignorance of the intervening fluctuations.

Secondly, the points are sometimes carelessly plotted so that though the general picture they give is correct, there is doubt about the detailed information that the chart provides. The difficulties are best made clear by asking the class to draw charts illustrating data of the following kind:

(*a*) Monthly totals.

(*b*) Stocks at end of month.

(*c*) Weekly averages per month.

Suitable data for exercises from which items of topical interest can be chosen can be found in any issue of the *Monthly Digest*

of Statistics (H.M.S.O. 2*s*. 6*d*.). When the diagrams are drawn it is an instructive exercise to redistribute them among the pupils and ask them to reconstruct the table of data on which the diagrams are based, or alternatively to read from the diagrams a few selected values.

2(*c*) *Trends, Seasonal Movements and Random Fluctuations in Time-Charts*

As time-charts are practically so important it may be worth spending two or three lessons on their analysis, particularly for pupils who are unlikely to continue their academic careers beyond the school-leaving age. The main ideas of the simple treatment of time-series analysis described below can be appreciated by children who have drawn only four or five charts, if the data are well chosen.

The feature of a statistical time-chart that is most obvious to a schoolboy is its irregularity. It may suggest broadly a simple pattern of recurrent peaks or a rather uncertain wandering up and down from one side of the chart to the other, but superimposed on the main trend there is usually a rather dazzling succession of random fluctuations. We need some simple method of separating the wood from the trees, i.e. of 'smoothing' the random irregularities so that the main features of the time-chart can be more readily seen and described.

The simplest way of showing the trend of an irregular chart is to draw a smooth curve 'through' the points, ignoring the minor fluctuations. This can be done visually but it takes a practised eye to do it well. The more sophisticated way of calculating a series of 'moving averages' is not difficult and is much more interesting. The averages of successive sets of readings (say 10) are calculated and are plotted on the graph so that each corresponds with the centre of the period spanned by the set. If the average of the first 10 readings $y_1, y_2, \ldots y_{10}$, is m_1, then the average m_2 of the next 10 readings, $y_2, y_3, \ldots y_{11}$, is given by

$$m_2 = m_1 + (y_{11} - y_1)/10$$

and so on.

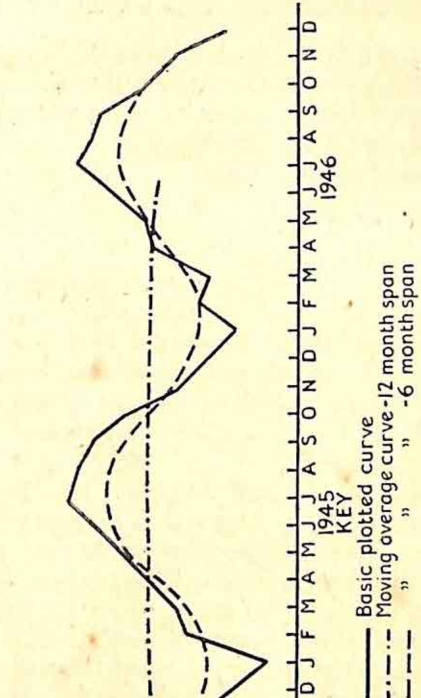

FIG. 3. Moving averages of cyclic data (mean daily temperature at sea-level; England and Wales).

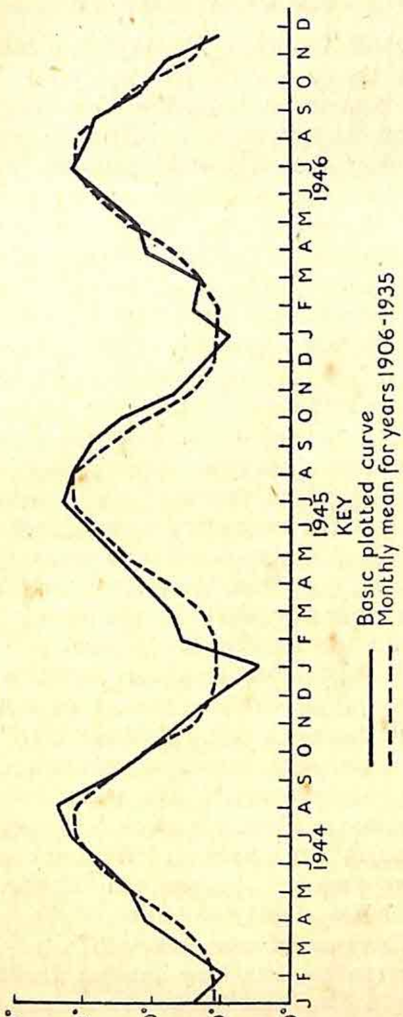

FIG. 4. Random fluctuations from the mean seasonal change.

If the time-chart shows a regular periodicity then some care in the choice of the span of the moving average is needed. If the periodicity recurs after 12 readings for example, as it does with any quantity that varies seasonally and of which monthly readings are taken, then sets of 12 (or some multiple of 12) readings must form the span of the moving average in order to smooth out the periodicity. If we wish to examine the periodic fluctuation itself then the averages of readings corresponding to successive phases of the period are calculated and plotted to show the average cycle.

All these points are illustrated in Figs. 3 and 4. Superimposed on the original chart are the charts of moving averages with 6-month and 12-month spans (Fig. 3). As we should expect, the 12-month span smooths out the seasonal fluctuations and shows that there is no appreciable change in mean temperature from year to year during the years considered. The curve based on the means for January, for February, etc., is also drawn (Fig. 4), and repeated from year to year; it shows the random fluctuations from the mean seasonal change, and enables us to say for any particular month whether or not it was 'hot' or 'cold' for 'the time of the year'.

Data suitable for illustrating these procedures with time-charts, if not readily available in the school, can be found in any issue of the *Annual Abstract of Statistics* (H.M.S.O. 10*s*.), but data of local or topical interest should be used if possible. Some suitable exercises are: 1. Examination of the daily absentee figures to see if there is any weekly periodicity (e.g. in schools at which attendance on Saturdays is required the figures may show a marked increase in the number of absentees on that day). 2. Day-to-day recording of the local meteorological data. These are of more interest if they are compared with the corresponding data from another locality, e.g. those published daily in *The Times* for London, or data provided by another school in a different part of the country. They should at least be compared with the known means for the locality with or without appropriate moving averages.

2(*d*) *Further Points about Charts*

Sometimes we may need to compare large numbers that are almost equal, where the important characteristic is not the approximate equality of the numbers but their differences. The bar-chart, as already described, may then fail to illustrate the characteristic in which we are interested. Is it reasonable to enlarge the scale of the bar-chart and chop off some of the inconvenient length? Yes, as long as we make clear what has been done. An alternative method is sometimes possible. If the large numbers were, for example, the number of pupils attending school, it might be more instructive to consider the number of absentees instead of the number of those present.

Finally, any chart should be as far as possible complete and self-explanatory. The title and the legends should be brief but adequate; any scale of numbers should be clearly marked.

2(*e*) *Maps*

The regional distributions of particular economic activities are very conveniently summarised by using maps of the region marked in a suitable way. Children are more likely to be interested in such maps if they have had opportunities of making them. Many good examples will be found in modern books on economic geography (e.g. *English County—A Planning Survey of Herefordshire*. Faber). With maps as with charts the information must be simply and fairly conveyed; false impressions are given, for example, by maps showing the results of parliamentary or local government elections by shading *areas*. A suitable class exercise is to prepare a map of the school's locality to illustrate 'The Journey to School'. The pupils' homes could be marked on the appropriate places by dots (say 1 dot for 5 pupils); their various methods of travel (bus, train, bicycle, etc.) could be indicated by streams of different colours; their routes by streams of width proportional to their traffic density. The collection and summarising of the necessary data is in itself a useful exercise; inevitably there will be cases which defy classification into any simple scheme

and which will demand more careful definition of the classes. The data can be used in later exercises on sampling.

To maintain the children's interest in diagrammatic representation some space on the classroom wall should be given to well-drawn charts illustrating subjects of topical interest, e.g. the local barometric data, sports results, house competitions, etc., and to a 'rogue's gallery' of bad and misleading charts brought by pupils.

3. Percentages, Rates and Ratios

A common defect in the teaching of percentages, rates and ratios, topics to which much time is usually given in school mathematical courses, is that too much emphasis is placed on the techniques of calculating them and too little on the reasons for their use, on their interpretation, and on the selection of the standard or base of the comparison. The techniques are comparatively simple but it seems to be too readily assumed that if children learn to know 'how' they will by some inner light learn also to know 'why'. That many intelligent children do not must be obvious if the frequent elementary mistakes and confusions to be met in everyday life are noted. Thus, to give one example, the B.B.C. recently announced that London taxi-drivers were on strike because their claim for an increase of 7% in their share of the metered fares they take had been refused by the taxi owners. What the drivers claimed in fact was an increase in their share from $33\frac{1}{2}\%$ to 40%, i.e. an increase of 20%—which perhaps more readily explains why a strike occurred. Here it is evident that confusion exists about the basis of comparison, though sometimes similar difficulties arise because the speaker or writer cannot express himself exactly even if he knows that precise expression is necessary.

Though the early teaching may be sound, many text-books of arithmetic contain a section of ingenious 'harder problems' which is well contrived to shatter for ever the confidence of most children in rates, ratios and percentages. It is suggested that the time spent on these trickier problems could be better

spent in discovering how to use rates, ratios and percentages in practical or social problems. A typical 'harder problem' reads: 'A man would gain 20% by selling an article for 9*s*. 6*d*. and 15% by selling a second article for 7*s*. 2*d*.; for what does he sell the second article if there is no loss or gain on the two sales?' For many children who have reached the stage (at 15 years) at which they are expected to solve this kind of puzzle a discussion of some measures of social importance, e.g. the cost-of-living index, would be of greater educational value. But even at an earlier stage an important point is commonly missed. Consider the simpler problem: 'On farm *A* 135 tons of wheat were harvested from 180 acres; on farm *B* 150 tons were harvested from 200 acres. Which farm had the greater yield per acre?' This problem gives little difficulty to many children who would be unable to answer in precise terms the question in this form: 'How could you find out which of two farms gave the greater yield of wheat per acre? What information would you need to know?' By putting the question in the numerical form it becomes a matter only of technique; the second form of the question requires a deeper understanding of the problem which children are usually expected to attain simply by repeated working of numerical questions.

Some problems needing judgment in the selection of appropriate measures of comparison are required in place of 'harder problems' or of some of the many problems which are merely straightforward applications of technique. Text-book and examination questions almost always specify the form in which the answer must be given. What percentage . . .? What is the proportion . . .? The advantage to the teacher and the examiner is that such questions have only one answer that is correct, so that marking is simplified, though much of the value of the problem to the pupil is lost. More examples of the following kind would be helpful:

1. Two football clubs *A* and *B* played a match which *A* won by 2 goals to 1. Would you approve or disapprove of the following alternative descriptions of the result:

(*a*) *A* won by 100%,

(b) A won by 50%,

(c) B lost by one goal,

(d) A was twice as good as B?

Give another way of stating the result which you think is suitable.

2. An article which cost 2*d*. in 1940 cost 6*d*. in 1950. State the increase in cost in three ways. Which is the most suitable way of describing this increase?

3. There are 20 pupils absent from school A and 15 pupils absent from school B. What further information would you need in order to make a fair comparison of absence from the two schools? Why?

4. Compound Units

Unless mechanics is taught in the school children rarely meet any of the compound units that are widely used in practical affairs. In fact some care to avoid them seems to be taken. A question from a modern text-book reads: 'A can do a piece of work in 30 days and B can do it in 6 days. How long will A and B take working together?' Intelligent children find such questions unrealistic. Though the teacher may see the question merely as a variant of the bath-tap problem and solve it automatically as such, many children would be conscious of the fact that two workers with such dissimilar capacities for work are unlikely to combine as harmoniously as the teacher assumes. Why are the practical units such as man-hours, ton-miles, kilowatt-hours, passenger-miles, acre-feet, etc., ignored? They are simple in principle, easy to manipulate, and their introduction to school courses would help to link the classroom with everyday affairs. Two suggested easy questions are as follows:

1. The re-surfacing of a road is estimated to take 4800 man-hours. The men work 8 hours per day; 30 men are available for 10 days and for the remainder of the time only 12 men are available. How many working days should the whole work take?

2. A wireless set consumes 80 watts when it is switched on. Estimate its cost per month of 30 days if it is used for an

average of $2\frac{1}{2}$ hours per day and if the supply charge is 2*d*. per unit (kilowatt-hour).

5. Averages

The average or arithmetic mean is the statistical measure most commonly used for making numerical comparisons. The principle of its computation is well understood but its limitations are rarely mentioned in school. Children see the average widely and indiscriminately used: they often see their marks in History, Art, Latin, Mathematics, Religious Knowledge, Science and Woodwork solemnly summed and averaged; minute differences in the result may serve to quell all possible disputes about the most worthy recipient of the form prize. Schools have long accepted the average as an infallible device for picking the winner. Out of school, the child sees in his newspaper the cricket averages calculated to two places of decimals and in ordered sequence; he reads that a difference of 0·001 in goal average may decide important questions of promotion and relegation in the football leagues. The average is only too evidently the final arbiter of all statistical problems, though unexpressed doubts may arise in boys' minds when they notice that Test Match selectors do not automatically choose the batsmen and bowlers with the best averages.

Four modest extensions of the usual limited teaching of the average are suggested for the general curriculum:

(*a*) That some consideration be given to the dispersion or 'spread' of numbers averaged.

(*b*) That it should be shown that the average is, for suitable data, merely a convenient shorthand description, useful for purposes of comparison, but that it is not an end in itself.

(*c*) That the computation of the average be taught more systematically.

(*d*) That commonly used applications of the average be more fully discussed.

Even without a quantitative measure of dispersion children can be shown, if their teacher takes his opportunities, that the average by itself is often inadequate as a description. Thus two

regions with approximately equal annual rainfall differ considerably in climate, scenery and economy if in one the rain falls only in the winter and in the other throughout the year; rivers that are either in flood or are dry are of much less use to man than a river which flows steadily throughout the year; manufacturers aim at uniformity in the quality of their products so that their customers can rely on getting reasonable value for their money; and so on. Sometimes the minimum or the maximum value is more important than the mean; the strength of a chain is determined by its weakest link; a bridge has to be designed so that it will safely withstand the estimated maximum stresses. Much can be done by comment of this kind whenever suitable numerical data are being used or discussed.

When an average is computed it is implicit in the operation, firstly, that the data used are homogeneous, and secondly, that the result is going to provide additional information that is useful. If, for example, a boy gets 5% for French and 95% for Mathematics, is it reasonable to add the marks together and is it helpful to know that his 'average' is 50%? It may of course be helpful to know that of two forms taking the same examination paper one form has an average of 60% and the other of 40%, but the calculation of the average as a matter of course or the averaging of data that are not homogeneous has nothing to recommend it. Discussion of the results of class experiments in the science laboratories can be used to illustrate some of the properties of the average and incidentally to help in the inculcating of the scientific attitude. Discrepancies between individual results, between the average result and the correct result (if known), can often throw light not only on individual errors but on deficiencies of technique. Small discrepancies, unaccountable at first, have often produced scientific results of great interest and importance (Lord Rayleigh's discovery of argon is a classic example) and children should be encouraged not to ignore them. Even the simple experiment of separating salt and sand by washing out the salt from a weighed sample followed by the drying and weighing of the sand residue produces interesting results. What are the

basic assumptions? That the sand is completely insoluble, that the salt is completely soluble in water, and that the washing water contains no suspended matter. Accepting these, what are the likeliest sources of experimental error? Insufficient washing out of the salt (How can we test the washings?), incomplete drying of the insoluble residue (How can we ensure that the residue is completely dried?). What effects would such errors have on the result of the experiment? What is the best way of stating the proportion of salt in the sample? Why are the results not all equal (assuming arithmetical errors to have been dealt with)? Was the original mixture homogeneous? Would it be sensible to *average* the result? What *is* the average result? Why is it different from the correct result (known from the composition of the mixture)? What is the range of the results? If we repeated the experiment again, avoiding all our known mistakes, what should happen to the range and the average of the results? Let us see if it does.

A more advanced experiment is typified by the determination of the internal resistance of a Daniell cell. The experiment is carried out in the usual way with no special care. What is the range of the results? Were repeated results equal? Why not? Does the internal resistance of a cell vary? Are the differences ascribable entirely to unknown experimental errors? Can we test this? Consider the cell; what would affect its internal resistance? Dirty terminals? Poor contacts? Dirty pots? Strength of copper sulphate solution or of nitric acid? Try again with clean terminals, fresh saturated copper sulphate solution, nitric acid of the specified strength, clean zinc rods, etc. Is the range of the results reduced? Discussion of the results is not merely a statistical exercise; it can teach much about laboratory physics.

As a chemical experiment consider the determination of the equivalent weight of copper. This is usually done by at least two methods, (*a*) by oxidising a known weight of copper to copper nitrate in a weighed test-tube and converting the nitrate to cupric oxide, (*b*) by reducing cupric oxide to copper in a stream of coal gas. Comparison of the means and ranges of the

two sets of results can show which method is the better and a discussion may explain why. To make the discussion realistic it must be conducted by the teacher who was in charge of the laboratory experiments—only he has the detailed background knowledge which is essential in interpreting the results.

In a simple course of human biology each member of the class can measure another member's 'vital lung capacity' (i.e. the volume of air breathed out during a deep exhalation). From these data the average capacity of all members of the class can be computed. Then for each member a series of five measurements at intervals can be taken. From these sets of data an average for each individual can be found. Finally, a series of measurements in quick succession can be made on one member of the class. As he tires his lung capacity becomes progressively smaller; the danger of accepting an average where its use disguises a trend can be demonstrated.

From such discussions some important statistical principles should begin to emerge:

(*a*) That any individual experimental result is usually subject to small random errors and should be regarded not as *the* result but as an estimate of the result.

(*b*) That the mean of a number of results (when the mean is appropriate) is still only an estimate, though one in which more confidence can be placed than in any individual result.

(*c*) That still more confidence can be placed in the mean result if, by careful technique, the spread of the individual results can be reduced.

(*d*) That differences between individual results, or between the mean and the correct result, are only to be expected; that they are not necessarily 'real' or 'significant', but arise by chance.

(*e*) That some differences are real or significant and must be explained in terms of some defect of technique or of some weakness in the basic assumptions (e.g. that the salt and sand were thoroughly mixed when in fact they were not).

A difficulty arises here—how is it possible for the teacher to decide whether a difference is 'significant' or not without

an elaborate statistical computation? A rough working rule for the teacher would be useful. If we assume that the distribution of results is approximately Normal with standard deviation σ it is known that the mean range, r, of samples of 4 results tends to the value $2{\cdot}06\sigma$ as the number of samples increases. By noting the ranges of the results taken in sets of 4 as they are worked out an estimate of σ is obtained by halving the mean range. If m is the calculated mean result and E the expected result, then the difference $d = | E - m |$ is 'significant', i.e. is highly improbable as a matter of chance if

$$\frac{d}{\frac{1}{2}r/\sqrt{4}} > 4, \quad \text{or} \quad d > r$$

For any individual result a deviation from the mean which is greater than $3r/2$ is also highly improbable as a chance result. It must be emphasised that these rules are only rough approximations intended to help the teacher and not to be given to the class: for a more elaborate analysis, which is necessary only as a matter of interest and as an occasional check, tables of the Normal probability function and of the t-test will be required.

For the more systematic computation of the average it is necessary to treat it not only as an arithmetical exercise in summation and division, but also as an exercise in algebra. There is nothing difficult in the use of Σ as shorthand for 'the sum of all things like . . .' and it can be introduced much earlier than with the summation of finite series, as is generally the case. Once the use of Σ has been explained by the use of some simple numerical examples, the arithmetic mean of x_1, x_2, x_3, . . . x_r . . . x_n can be defined as $\bar{x} = \frac{1}{n}\Sigma_1^n x_r$. Simple extensions of this formula which are useful in the rapid computing of averages, such as those required for the discussion of experimental results obtained from the laboratory, are:

(a) If $x_r = a + y_r$ etc., where a is a constant, then

$$\begin{aligned}\bar{x} = \frac{1}{n}\Sigma_1^n x_r &= \frac{1}{n}\Sigma_1^n (a + y_r) \\ &= \frac{1}{n}(\Sigma_1^n a + \Sigma_1^n y_r) \\ &= \frac{1}{n}(na + \Sigma_1^n y_r) \\ &= a + \frac{1}{n}\Sigma_1^n y_r\end{aligned}$$

e.g. the mean of 981·6, 980·5, 980·7 and 982·3 is equal to 980 + (the mean of 1·6, 0·5, 0·7 and 2·3).

(*b*) If $x_r = c . y_r$ etc., where c is a constant, then

$$\begin{aligned}\bar{x} = \frac{1}{n}\Sigma_1^n x_r &= \frac{1}{n}\Sigma_1^n c . y_r \\ &= \frac{c}{n}\Sigma_1^n y_r\end{aligned}$$

e.g. the mean of 925, 825, 875 and 975 is equal to 25 × (the mean of 37, 33, 35, and 39). This result is useful to introduce before frequency distributions are analysed.

(*c*) If $\bar{x}_1$ is the mean of a set of n_1 numbers, and $\bar{x}_2$ is the mean of a set of n_2 numbers, then the mean of the set obtained by combining the two sets of n_1 and n_2 numbers is given by

$$\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

(*d*) The above formula has many and diverse applications. Expressed in the more general form as

$$\bar{x}_w = \Sigma_1^n w_r x_r / \Sigma_1^n w_r$$

it is known as the *weighted mean*. The 'weights', w_1, w_2, w_3 . . . w_n, or their ratios, are usually based on statistical data but, as in the 'weighting' of examination results, they may be arbitrarily assigned. In mechanics it is used in problems involving parallel forces or the dynamics of a number of particles; in co-ordinate geometry one of its uses is to give the co-ordinates of a point which divides a line in a given ratio; in economic

statistics it is of importance in constructing compound index numbers. Because of its wide applicability it deserves greater attention in elementary mathematics than it receives at present.

Some examples:

1. The average wages of men and women in an industry are 107*s*. and 92*s*. per week respectively. The ratio of men to women employed in the industry is 3 : 2. What is the average wage for all workers?

2. Iron is a mixture of two isotopes of atomic weight 54 and 56. The atomic weight of the mixture is 55·84. What is the composition of iron in terms of its isotopes?

3. 500 c.c. of a 2N solution of hydrochloric acid are added in error to a bottle containing 7500 c.c. of a 0·1N solution of the same acid. What is the strength of the mixture?

4. Metallic ores containing 25%, 21%, and 18% of the metal are used in the ratios 3 : 5 : 2 respectively. What is the content of the metal in the mixture?

5. In a certain industry the average weekly wages of men, women and juveniles are 112*s*., 85*s*. and 48*s*. and the ratios of the numbers of each employed are 6 : 3 : 1 respectively. What is the average weekly wage for all the workers of the industry?

6. The average humus content of a sandy soil is 11·2%, and that of a peaty soil is 53·8%. In what proportions should the two soils be mixed to provide a soil with a humus content of 25·4%?

A further property of the mean that is important is that the algebraic sum of the deviations from the mean is zero, i.e. $\Sigma_1^n(\bar{x} - x_r) = 0$. If elementary mechanics is taught it is instructive to find the mean of a few numbers practically by loading a light beam and finding the point of balance, e.g. to find the mean of 66, 92, 5, 12, 70 and 55, suspend 1 lb. weights at 66 cm., 92 cm., etc., from one end. The point of balance in this example is at 50 cm., so that the weight of the beam does not affect the result, but small deviations of the mean from the centre of the beam will not seriously affect the result. The metre rule can also be used to demonstrate the weighted mean.

6. Index Numbers

Though there are many practical difficulties in the computation of index numbers the ideas on which they are based, percentages and weighted means, are well within the grasp of 14–15-year-old children. The simplest type of index number is that which measures the price fluctuation of a single commodity. The price of a commodity is noted over a period and a norm is arbitrarily selected; it may for example be the price of the commodity at a given date, or its average price over a comparatively stable period. The price is then used as a standard against which later prices are compared; the later prices are expressed as percentages of the norm. Changes in the index number are easier to appreciate than changes in the cash prices of the commodity, and the index numbers of different commodities are easier to compare than the cash prices of those commodities. Thus for strawberries sold in England and Wales and with the average price for 1936/38 as 100:

Year . .	1936	1937	1938	1939	1940	1941
Price per lb.	9*d.*	9*d.*	10¼*d.*	9¼*d.*	1*s.* 0¾*d.*	1*s.* 7¾*d.*
Index . .	95	95	109	98	135	210

Year	1942	1943	1944	1945	1946
Price per lb. . .	1*s.* 3*d.*	1*s.* 3½*d.*	1*s.* 2¾*d.*	1*s.* 4*d.*	1*s.* 2½*d.*
Index	159	165	156	170	154

Similar simple index numbers for other fruits are made and from these data a composite index number for all fruits is computed as a weighted mean. In the same way composite index numbers for vegetables and glasshouse products, for livestock and livestock products, and for cereals and farm crops are computed from their several items. The mean of the composite index numbers, suitably weighted, yields the 'agricultural price index'.

The calculation of a few composite index numbers from the simple index numbers provides good exercises in the use of the weighted mean. The *Monthly Digest of Statistics* (H.M.S.O.

2*s*. 6*d*.) and the *Monthly Bulletin of Statistics* published by the Statistical Office of the United Nations (H.M.S.O. 2*s*. 6*d*.) and their occasional supplementary volumes provide much suitable data of topical interest and information about index numbers. At the same time the exercises provide opportunities of mentioning some of the uses and abuses of index numbers and of removing some misconceptions. A discussion of cost-of-living index numbers is worth while, if the following points are emphasised:

(*a*) Cost-of-living index numbers of different countries are computed in different ways to serve different purposes and are not strictly comparable.

(*b*) They do not indicate whether it costs more to live in one country than in another.

(*c*) They need to be modified from time to time to reflect changes of taste, habits and conditions (as was done in June 1947 in the United Kingdom).

(*d*) They are averages based usually on samples of city working-class expenditure; large deviations may occur in individual cases and from one region of a country to another.

For more advanced mathematical work two additional items might be considered:

(*a*) The weighted geometric mean as an index; its advantages and its computation.

(*b*) The change in a composite index number caused by small changes in its components.

7. Measure of Dispersion

The introduction at this stage of any measure of dispersion other than the range is hardly justifiable, unless time is available for a first study of frequency distributions. The danger of introducing the standard deviation (or even the mean deviation from the mean) too early is that it may be learnt (and taught) merely as a trick of technique. For many classroom purposes the simple range is adequate, but its main defect as a measure of dispersion is that it depends directly upon only two values of the variable. A more efficient way of using it is to find the

mean range of a number of small samples of equal size as has already been described (p. 19 above). The samples should be selected in some random way from the data available, and, if possible, at least 6 samples should be taken. As the mean range of the sample varies with the sample size it is important that equal sample sizes should be used when the two dispersions are being compared. As an alternative (as an aid to the teacher but not for classroom use) the factors of Table 2 may be helpful; a rough estimate of the standard deviation of the distribution is given by multiplying the mean sample range by the given factor.

TABLE 2

Factors for converting mean ranges to rough estimates of the standard deviation

Sample Size .	2	3	4	5, 6, 7	8, 9, 10
Factor . .	0·9	0·6	0·5	0·4	0·33

8. Frequency Distributions

When we are confronted by a large number of observational results it may be difficult to grasp the main features of the data, that is, to 'see the wood for the trees'. A frequency distribution is an orderly arrangement of the data. For example, a chemistry master suspected that his class of 14-year-old boys were being careless in their readings of the burettes they were using in an experiment and the use of which had already been explained to them. To check their accuracy several burettes were set up with different levels of liquid and the boys were asked to move from one burette to the next, noting and recording their readings. The following is a record of the readings set down by 25 boys for a burette which, according to the chemistry master, had a correct reading of 23·31 c.c.:

23·35	23·35	23·30	23·30	23·35
23·30	23·30	23·40	23·35	23·32
23·38	23·30	23·35	23·30	23·30
23·35	23·32	23·40	23·35	23·30
23·35	23·32	23·35	23·30	23·30

At a glance this table of results conveys very little apart from the fact that the readings are unequal. To see more clearly what is happening it helps to arrange the results as a 'frequency distribution':

Reading .	23·25	23·30	23·32	23·35	23·38	23·40	Total
No. of boys	2	10	3	7	1	2	25

With the help of a diagram of the liquid meniscus in the burette tube it was possible to show both the faults of individuals and the faults of the class as a whole. The main mistakes can be seen to be:

(*a*) Insufficient care in positioning the eye.

(*b*) Rounding off to the nearest unit or half unit.

(*c*) Possible miscounting of the unnumbered divisions of 0·1 c.c.

Such informal introductions to frequency distributions, especially if they are related to some work of the class, will simultaneously explain what they are and how they are used, before they are considered in a more formal way.

The next step is to introduce frequency distributions of grouped data and their representation by histograms. For this purpose examination marks on the scale 0–100 are suitable, but at least 100 marks should be available. The marks are read out and plotted, without grouping, one by one. The resulting chart will show the features of the distribution in such detail that it is not easy to describe it simply in a general way. A less detailed chart might be more effective—so let us group the marks in equal class intervals of 5 marks, 1–5, 6–10, . . . 96–100. (If any 0's are to be accounted for, it is better for the present to include them in the class 1–5 to avoid unequal class intervals at this stage.) The results are tabulated:

Mark	*Frequency*
1– 5	3
6–10	7
11–16	12

etc.

and the 'histogram' is drawn, as though it were a bar-chart, with columns of height proportional to the corresponding frequencies as in Fig. 5. There is no particular reason for selecting a class interval of 5 marks, and as it is interesting to compare the histograms obtained by using different class intervals, the members of the class should be allowed some choice from 2, 4, 5, 10, 20 and 25 marks, which all give sets of equal class intervals for the range 1–100. Discussion of the results should show that a total of about 20 classes displays the results in sufficient detail for most purposes: if the class intervals are too large, important features may be masked.

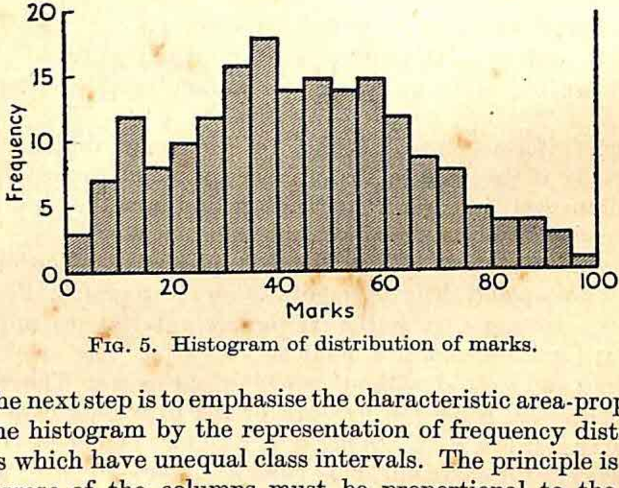

FIG. 5. Histogram of distribution of marks.

The next step is to emphasise the characteristic area-property of the histogram by the representation of frequency distributions which have unequal class intervals. The principle is that the *areas* of the columns must be proportional to the corresponding frequencies; the *heights* of the columns are proportional to the frequencies only for the special case of equal class intervals. Suitable data can be found in the *Annual Abstract* and can be chosen to illustrate histograms of different shapes, e.g. V-shaped, J-shaped and skew. This area-property of histograms is of great importance for later work in sampling and probability theory.

An example with the calculations is set out in Table 3. The

data illustrate some striking changes which are best brought out by drawing the two histograms on the same diagram. A difficulty arises about the last interval which is 'open'; unless more facts are given only a judicious guess can be made about it.

TABLE 3

Deaths in England and Wales; a comparison of age at death in 1871 *and* 1946

Age	1871	1946	Column dimensions		
			Base	Heights	
				1871	1946
0–1	126	33	1	126	33
1–2	41	2	1	41	2
2–4 +	40	3	3	13·3	1·0
5–14 +	31	5	10	3·1	0·5
15–24 +	30	8	10	3·0	0·8
25–34 +	34	13	10	3·4	1·3
35–44 +	33	20	10	3·3	2·0
45–54 +	35	39	10	3·5	3·9
55–64 +	42	76	10	4·2	7·6
65–74 +	48	127	10	4·8	12·7
75 +	49	162	?	?	?

A rather more sophisticated exercise is to draw the histogram of a skew distribution, first on a linear base and then on a logarithmic base or, alternatively, on semi-logarithmic paper. Data suitable for this exercise will be found in the *Annual Abstract*, e.g. distribution of incomes liable to surtax.

The computation of the mean of a frequency distribution is shown in Table 4. This tabulated computation is easy to teach as a technical trick; it should be omitted if it is unlikely that its underlying principles will be understood. At this stage only distributions with equal class intervals and not more than 10 classes should be used. For this tabulated work squared paper ($\frac{1}{4}$-inch) is very helpful in keeping the work tidily arranged.

TABLE 4

Computation of the Arithmetic Mean

Mark	Frequency (f)	Centre of interval [1] (x)	Deviation from working mean		$f \times d$
			Marks [3]	Class intervals (d)	
1–10	3	$5\frac{1}{2}$	$-$ 40	$-$ 4	$-$ 12
11–20	8	$15\frac{1}{2}$	$-$ 30	$-$ 3	$-$ 24
21–30	12	$25\frac{1}{2}$	$-$ 20	$-$ 2	$-$ 24
31–40	15	$35\frac{1}{2}$	$-$ 10	$-$ 1	$-$ 15
41–50	16	$45\frac{1}{2}$	0 [2]	0	$-$ 75
51–60	20	$55\frac{1}{2}$	+ 10	+ 1	+ 20
61–70	12	$65\frac{1}{2}$	+ 20	+ 2	+ 24
71–80	9	$75\frac{1}{2}$	+ 30	+ 3	+ 27
81–90	4	$85\frac{1}{2}$	+ 40	+ 4	+ 16
91–100	1	$95\frac{1}{2}$	+ 50	+ 5	+ 5
	100				+ 92
					Total + 17 [4]

NOTES (see corresponding reference numbers in Table 4).

1. If this column is omitted in the first examples many mistakes arise. The entries in this column stress the fact that by grouping into classes we are placing all the marks of each class at the centre of that class. It is a useful exercise to show by examples that this approximation leads to negligible error if the curve is roughly symmetrical.

2. The working mean is chosen as a guess at the centre of the class which appears most likely to contain the mean. If the guess is wrong it does not matter, but the more nearly correct it is the less computing work there is to do.

3. This column can also be omitted after the first few exercises. It acts as a reminder that the units of the computation are being changed for the last two columns.

4. The mean is given algebraically by

$$\bar{x} = x_0 + C.\frac{\Sigma fd}{\Sigma f}$$

where x_0 = working mean
and C = class interval

The expression $\Sigma fd/\Sigma f$ should be recognised as a weighted mean in which the weights are the 'frequencies' of the second column.

The computation is completed thus:

$$\begin{aligned}\text{Mean} &= \text{working mean} + \frac{17}{100} \text{ class intervals}\\ &= \quad ,, \qquad ,, \quad + \frac{17}{100} \times 10 \text{ marks}\\ &= (45{\cdot}5 + 1{\cdot}7) \text{ marks}\\ &= 47{\cdot}2 \text{ marks.}\end{aligned}$$

The first few exercises should be arranged so that some give exact values of the mean; it is then possible and important to verify that 'the sum of the deviations from the mean is zero' exactly, i.e. that

$$\Sigma f(x - \bar{x}) = 0$$

As a measure of dispersion the standard deviation is to be preferred, but defined as 'the square root of the mean of the squared deviations from the mean' it sounds rather formidable to 15-year-old children. Though its computation requires the addition of only one more column to those of Table 4, there is no point in using it unless it can be thoroughly comprehended. For this stage therefore the mean deviation from the mean is probably the simplest measure to use. Unless the error between the working mean and the true mean is greater than half a class interval the mean deviation can be found with sufficient accuracy directly from the tabulated computation of the mean.

Thus for the data of Table 4:

$$\begin{aligned}\text{Mean deviation} &= (75 + 92)/100\\ &= 1{\cdot}67 \text{ class intervals}\\ &= 16{\cdot}7 \text{ marks.}\end{aligned}$$

Neither the computation of the mean nor of a measure of dispersion is an end in itself, except possibly for the one or two examples required to illustrate the technique of computing them; they should be computed only when they are to be used in comparing the properties of two or more frequency distributions.

9. Correlation

Almost all graphs drawn in school are those of simple mathematical functions or of experimental results which can be expected to give good straight lines or simple smooth curves. It is therefore instructive to consider some variables which are not so exactly related, e.g. the heights and weights of the members of the class, or their marks in, say, mathematics and physics. If such bivariate distributions are plotted on graph paper, rather dispersed but not entirely featureless patterns of points will be obtained. Though it may not be possible to draw at sight one straight line which can adequately represent the association between the two variables, yet it may be possible to say that, on the whole, high and low values of one variable are frequently combined with high and low values respectively of the other. Such distributions are of great importance in statistics; a simplified analysis by graphical methods as described below is possible for 15-year-old children.

When two variables x and y are plotted on a graph in the usual way, one of many possible patterns will result, though it is possible to arrange the patterns into a few typical groups. Six types of pattern are shown in Fig. 6. The first, Fig. 6 (a), is the graph of an algebraic relation of the form $y = mx + c$ where m and c are constants. Every pair of values of x and y lies *exactly* on the straight line. Lines as exact as this are mathematical ideals which, though important in geometry, are never realised in plotting practical measurements. Fig. 6 (b) illustrates the plotting of results obtained in careful experiments when some linear relation between two physical variables is being investigated. The plotted points lie in a narrow band about an ideal straight line which is usually drawn 'by

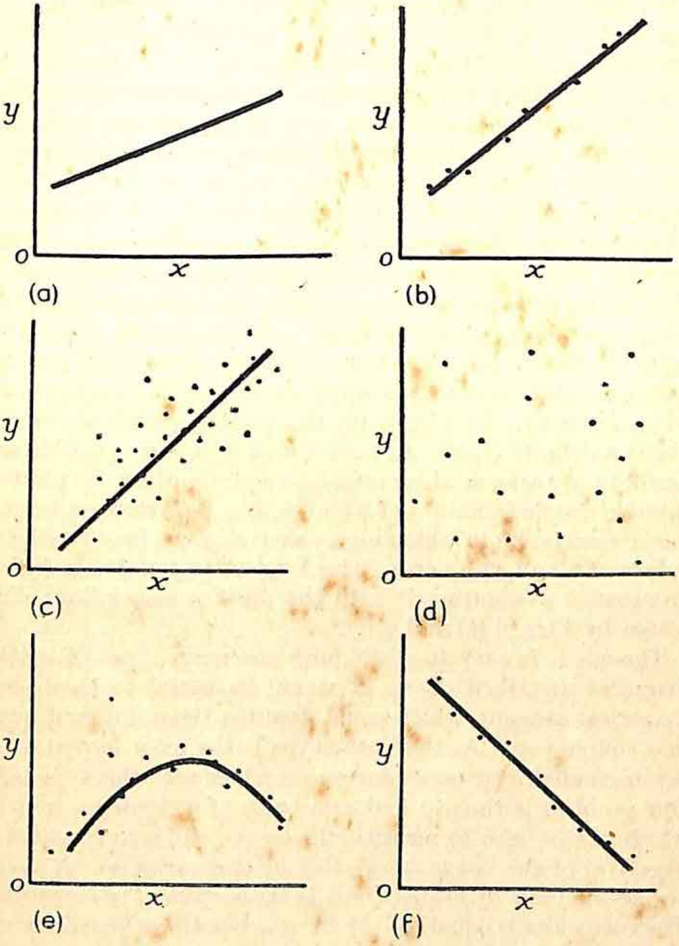

FIG. 6. Six types of two-variate association:—(*a*) Exact linear (mathematical); (*b*) Linear association (e.g. practical physics); (*c*) Linear association (e.g. practical biology); (*d*) No association; (*e*) Curvilinear association; (*f*) Negative association.

eye'. The plotting of pairs of biological measurements, e.g. the age and height of the members of the class, gives a greater scatter of points. The pattern shown in Fig. 6 (*c*) is of this type; though the points are widely scattered it can be said that, roughly speaking, high values of x are associated with high values of y, and low with low. In such a case there is said to be *some* linear association between x and y. As the degree of association *increases* the pattern becomes more like that of Fig. 6 (*b*): as the degree of association *decreases* the pattern becomes more like that of Fig. 6 (*d*) in which there is no discernible association between x and y; any selected value of x (or y) may equally well be found with any value of y (or x), high, medium or low. The relation between two closely related variables may not always approximate to the straight line form, however. In Fig. 6 (*e*) the plotted points lie closely about a definite curve. In such a case it is often possible and useful to 'transform' this curve into a straight line by plotting suitable simple functions of x and y. Fig. 6 (*f*) shows a form of linear association in which high values of y are found with low values of x and vice versa. This kind of association is said to be *negative* to contrast it with the *positive* associations illustrated by Figs. 6 (*b*) and 6 (*c*).

Though it is easy to distinguish the main types of scatter diagrams described above, it would be useful to have some numerical measure which could describe these different types in a simple way. As the linear type is the most important in practical affairs we need concern ourselves with this type only. Our problem is thus to find some way of assigning a number which can be used to measure the degree and sign (positive or negative) of the linear association of two variables. A simple coefficient used by statisticians is the *coefficient of correlation*. The coefficient is equal to $+ 1$ for exact positive linear association; it is equal to $- 1$ for exact negative linear association; and it is zero if there is no association at all. For intermediate degrees of linear association it can take any value between $+ 1$ and 0, or between 0 and $- 1$. This coefficient is defined in mathematical terms. Its calculation however is rather labori-

ous and for many cases in which it is useful an approximate value of it can be obtained graphically.

Assume that we have a set of corresponding mathematics and physics marks available, both sets in the scale 0–100. The graphical procedure is:

(*a*) Plot the pairs of marks on graph paper.

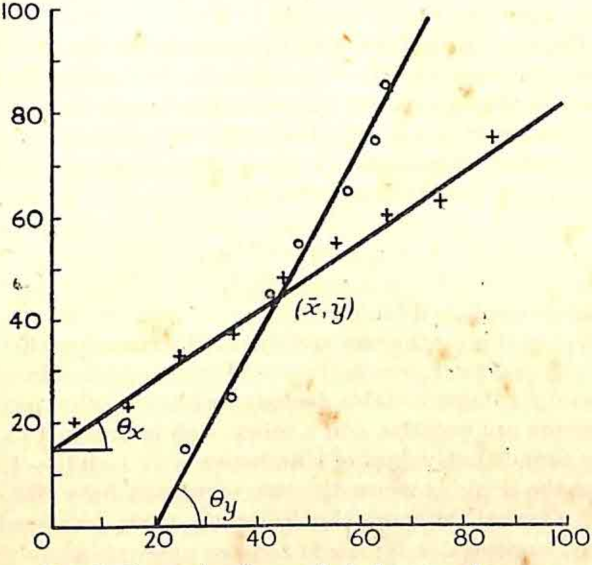

FIG. 7. Regression lines as loci of means of arrays.

(*b*) Find the means of the mathematics marks, $(\bar{x})$, and of the physics marks, $(\bar{y})$. Plot the point $(\bar{x}, \bar{y})$ (Fig. 7).

(*c*) Consider the marks in the class intervals 1–10, 11–20, 21–30, etc. For the mathematics marks 11–20, for example, there is a distribution of physics marks; plot the mean of this distribution by a cross on the line corresponding to the mathematics mark $15\frac{1}{2}$. The mean may be calculated but it is usually possible to estimate it sufficiently accurately by eye. Repeat this plotting of the means for all the columns (Fig. 7).

(*d*) Draw through $(\bar{x}, \bar{y})$ a straight line passing as nearly as possible 'through' the means of the columns. This line is called the line of regression of y (physics marks) on x (mathematics marks) (Fig. 7).

(*e*) Similarly for each row $(1 - 10, 11 - 20, 21 - 30$, etc., of physics marks) plot on its centre line the mean of the distribution of mathematics marks in that row. Use a small circle to indicate these means.

(*f*) Through $(\bar{x}, \bar{y})$ draw a line passing as nearly as possible 'through' the means of the rows. This line is the line of regression of $\bar{x}$ (mathematics marks) on $\bar{y}$ (physics marks).

(*g*) Find the slopes θ_x, θ_y of the two regression lines. From these slopes an approximate value of the coefficient of correlation, r, can be found. It is given by

$$r = \sqrt{\frac{\tan \theta_x}{\tan \theta_y}}$$

for reasons explained later.

If the trend is for the two variables to increase together both the $\tan \theta_x$ and $\tan \theta_y$ are positive and the positive value of the root is taken; if one variable decreases while the other increases both slopes are negative and a minus sign is assigned to r. It will be found that values of r lie between $+ 1$ and $- 1$. The smaller the angle between the two regression lines (for equal x- and y-scales) the more closely are the variables correlated, positively or negatively; if r is zero, or approximately so, the variables are said to be uncorrelated, and the angle between the two regression lines is almost a right angle.

When interpreting the correlation coefficient it must be remembered that a high value of r does not imply that there is necessarily a *causal* relation between the two variables; a low value of r will certainly be obtained if the variables are completely independent, but it may also arise if the variables are related in some way other than linearly, e.g. parabolically.

As the graphical method described above is applicable only to large samples (e.g. at least 100 pairs) sets of marks are likely to provide the most interesting ready-made data for class use.

Questions about the significance of the correlation coefficient must be deferred, and it would be wise to use in exercises only data for which $|r| > 0{\cdot}3$. (The 5% significance level of r for 100 pairs is 0·2 approximately.)

Children are always surprised to find that the graphical method described produces two lines instead of one, and some explanation of this phenomenon will be demanded. A bivariate distribution can usually be considered from two points of view, what we may call the 'x' and the 'y' points of view. The mathematics master considering the pairs of marks used in our example will have the x point of view; he will ask himself 'How do the physics marks compare with the mathematics marks?' In other words, he will use the mathematics marks as the basis of comparison and consider how the physics marks vary for given marks in mathematics. The line of regression of y on x summarises the distribution for him. The physics master, on the other hand, will consider the marks from the y point of view: 'How do the mathematics marks compare with the physics marks?' For him the line of regression of x on y summarises the distribution. The two lines are therefore an expression of the fact that there are at least two different ways of considering the marks.

The coefficient of correlation is a means of combining these two points of view into one expression about which both masters can agree; it is the geometric mean of the slopes of the two lines of regression though it may not appear to be so in our formula. The reason is that the line of regression of y on x is usually expressed in the form $y = ax + c$ so that $a = \tan\theta_x$, but the line of regression of x on y is expressed in the form $y = bx + d$ so that the slope b must be measured from the 'y' point of view. Hence

$$b = \tan(90° - \theta_y) = \cot\theta_y = 1/\tan\theta_y$$

$$r\sqrt{ab} = \sqrt{\tan\theta_x/\tan\theta_y}$$

The kind of data readily available in schools which is suitable for exercises in correlation depends very much on the kind of work done in the school. If the school has well-equipped

laboratories, or a school garden with an experimental plot, or a meteorological station, almost every class experiment provides some data worth a closer analysis than it is usually given. It is not suggested that more weighing and measuring be done (there is perhaps too much of this already) but that the results of some class experiments should be analysed as a whole rather than be scattered and lost as 30 or 40 isolated results separately recorded in individual notebooks. For example, in the elementary chemistry experiments in which a weighed quantity of a substance loses weight on heating, the class results can be plotted on graph paper, weight heated (x) against weight remaining (y), so long as the weight heated has a fair range of values. By this means the teacher of chemistry avoids some otherwise inevitable teaching of arithmetic and yet gains more than he loses in trying to explain the quantitative facts, while each pupil benefits by sharing his results with others. An exactly similar procedure may be followed in practical biology when the class is estimating the 'loosely-held water' (i.e. the percentage loss of weight on air-drying) in samples of the same soil specimen.

Where laboratory facilities are limited, the weight of objects up to 1 lb. (or 500 gm.) can be quickly determined to within ¼ oz. (or 10 gm.) by using a spring or lever letter balance. The measurement of lengths up to 12 inches (or 30 cm.) to the nearest 0·1 inch (or 1 mm.) should afford no difficulties. Many easy quantitative experiments on rate of growth, effect of fertilisers and hormones, germination, etc., requiring only simple apparatus, are described in *Simple Experiments in Biology* by Cyril Bibby (Heinemann, 8*s*. 6*d*.).*

An occasional spurious correlation can be both amusing and instructive.

10. Probability and Simple Probability Calculations

It has been argued that the subject of probability is unfit to be mentioned in school because it is still beset by logical and

* See also H. Kalmus, *Simple Experiments with Insects* (Heinemann, 7*s*. 6*d*.).

philosophical difficulties. If this argument were generally applied to the subject matter of the school curriculum very little of that curriculum would remain. Primary school courses of arithmetic do not begin with a study of *Principia Mathematica*, nor is it necessary to begin the study of probability with a critical examination of its possible definitions. Education at the school age rests mainly on the child's intuitive acceptance of ideas rather than on their logical necessity. If it is accepted that a more critical approach can be made only when the fine logical distinctions are appreciated, there is no reason why an introduction to probability should not be given at a much earlier age than it is at present. If the subject of probability is taught in schools at all it is usually as an unattached addendum to the Sixth Form algebra course for mathematical specialists, and then it is restricted to a brief treatment of the classical theory. In the suggested treatment of probability outlined below an attempt is made to relate the child's experience of everyday affairs to a subject which has many accepted practical applications.

Why do captains of football and cricket depend on the toss of a coin to decide who shall play the first half with the wind or take first innings on the batsman's wicket? It seems to be generally accepted as the fairest way of beginning a game. But what do we mean by 'fair'? Let us look at the coin more closely. Would it be fair to use a coin with two heads? or one that is bevelled or weighted to give heads more often than tails? No, we want to give each side an equal chance. What, then, do we expect of the coin? What are we taking for granted when we use a coin to toss for first innings? We are assuming that, if the coin were tossed a large number of times, it would give about as many heads as tails. If we tossed a penny 200 times would we expect it to give exactly 100 heads and 100 tails? No, it might happen to do so but we should not suspect the penny of being unfair or biassed if it gave 101 heads, or 105 heads, or 110 heads—but here we begin to doubt. (It is not necessary to decide where the line should be drawn, so long as the points are established that the number of heads

and tails need not be *exactly* equal in a finite number of throws, and that we do not expect an exact alternation of heads and tails in successive throws.)

It is now time to express in mathematical terms the fact that a fair coin will give, 'in the long run', about as many heads as tails. We say, mathematically, of an event that is impossible that its probability is 0, of an event that is certain that its probability is 1. What can we say about the probability of getting a head or a tail with a fair coin in a fair throw? The probability of getting one or the other is 1, because a coin dropped on a flat floor will fall with one side uppermost; the probability of getting neither is 0, because a penny thrown into the air does not remain spinning in the air indefinitely. We have to say that the probabilities of a head or of a tail are equal and that together they total 1; they must each be $\frac{1}{2}$. If the coin were biassed so that it gave, in the long run, more heads than tails, say 6 heads for every 4 tails, we would say that the probability of a head is 0·6, and of a tail 0·4.

The joy of playing snakes and ladders, ludo, and other games with dice is that one never knows what is going to happen next; with every throw we put ourselves at the mercy of chance. Let us look at a die more closely. It is approximately cubical, with its six sides numbered 1 to 6. What do we expect of a fair die? If we threw it many times, we should expect it to give roughly equal numbers of 1's, 2's, 3's, 4's, 5's and 6's, i.e. there should be equal chances of getting each of the numbers. As the probability of getting some number in a throw is 1, the probability of getting a particular number with a fair die is $\frac{1}{6}$.

Could the two captains use a die instead of a coin to decide whose team should bat first? They need to make its use as simple as possible, i.e. one throw, and yet they need to be fair to both captains. If the die is unbiassed, what about the number of odd and even numbers thrown in the long run? They will be roughly equal and therefore the probability of throwing an even number or an odd one is $\frac{1}{2}$. The 'toss' could therefore be decided by calling odds or evens with a die instead of heads

or tails with a coin. Notice that we can get the same result by saying that the chance of throwing either a 1, a 3, or a 5 is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. What then are the probabilities of throwing (*a*) a 6, (*b*) any number except 6, (*c*) a number greater than 4, (*d*) a number less than 5 ?

If the two captains had neither a coin nor a die could they use a pack of cards to decide the 'toss'? What should we like to know about the cards? Is it necessary to have a complete pack? etc. What is the probability of drawing from a shuffled pack (*a*) the ace of spades, (*b*) the two of diamonds, (*c*) a heart, etc.?

At this stage a class experiment with dice could well be undertaken, e.g. the comparison of the results obtained by throwing a good die and another die which has been biassed (e.g. by filing down one face). Two pupils can alternate in throwing the die and announcing the result while the remainder of the class individually record the results on prepared graph paper. The significance of the phrase 'in the long run' becomes clearer if, at convenient intervals, the proportionate deviations of the observed results from those expected are compared.

So far only the probability of occurrence of single events has been considered. An introductory treatment of compound probability, i.e. the probability of occurrence of two or more independent events, is possible.

What is the probability of winning the toss on two successive occasions? To answer this let us begin by listing all the possible results of tossing a coin twice, namely, (H H), (H T), (T H), (T T). All four possibilities are equally likely so the chance of calling any one of them correctly is $\frac{1}{4}$. The probability of calling correctly in three successive throws can be investigated in the same way but unless the binomial theorem is already known it is not possible to generalise the probability distributions of heads and tails. The following question on the topic is sometimes asked: If a penny has given 100 heads in 100 tosses what is the probability that the next toss will also give a head? The answer usually expected is $\frac{1}{2}$, but a penny that behaves in such a way is almost certainly biassed. With-

out further information one can only say that the probability of another head is much nearer to 1 than to ½.

The probable results of throwing two dice are interesting and to children rather surprising. With one die there is an equal chance of throwing any of the numbers from 1 to 6. With two

Die A

Die B		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

FIG. 8. Combination of two dice scores.

dice we throw any of the numbers from 2 to 12; are they equally probable? A chess-board scheme displays the possibilities in the clearest way (Fig. 8). All the arrangements shown are equally likely and therefore the 'probability distribution' of the throws of two dice is:

Sum of the two dice	2	3	4	5	6	7
Probability . . .	1/36	2/36	3/36	4/36	5/36	6/36
Sum of the two dice . .		8	9	10	11	12
Probability		5/36	4/36	3/36	2/36	1/36

The probability of getting 7 in a double throw is therefore 6 times as great as the probability of getting 12. These and

similar results which arise from the drawing of cards from sets numbered 1–10 (e.g. playing cards) are better appreciated if the expected results are compared with some experimental results. In using cards, those that are drawn must be replaced and carefully reshuffled between each draw to make the drawing process exactly analogous to the throwing of dice. Cards soon tend to stick in pairs; dice are the easier and quicker to use.

There is another aspect of probability which will intrigue children who like noting registration numbers of cars and 'logging' numbers of railway engines. Here is a game. One-tenth of all registration numbers end in the same digit, let us say 5. Would you expect precisely one last digit of 5 in every ten cars? No. What would you expect? Ask the children to find out by actually counting tens of motor-cars and scoring the number in each ten which have the selected final digit. The distribution of a large number of such scores will be found (if the observers are honest) to follow a regular pattern, though of course not exactly. The regular pattern is given by the 'probability distribution' (which in this case is the Poisson distribution with a mean equal to 1) shown in Table 5.

TABLE 5

Probability distribution of car registration numbers ending in a given digit in 100 *random groups of* 10 *cars*

No. in group	0	1	2	3	4	5 or more	Total
Frequency .	37	37	18	6	2	0	100

It is usually surprising to children to find that so 'chancy' a game reduces, in the long run, towards a mathematical regularity.

A variant of this game of observing the last digit of a car registration number is to select a final digit and then count the number of cars that pass until the digit recurs. The number is scored as 1 if the next car to pass is a 'success', as 2 if the first car is a 'failure' and the second car is a 'success', and so

on. The distribution of these scores, in the long run, becomes a geometric one. The probabilities of scores of 1, 2, 3, 4, . . . are $\frac{1}{10}$, $\frac{1}{10}(\frac{9}{10})$, $\frac{1}{10}(\frac{9}{10})^2$, $\frac{1}{10}(\frac{9}{10})^3$ . . . respectively. It is perhaps surprising to find that the most probable score is 1, though the average score is, theoretically, 10.

These simple games effectively illustrate how chance, usually considered to be the negation of law, is itself in the long run constrained by laws. The examples themselves are of course trivial and of no practical importance, but they illustrate laws of probability that are applied to many practical problems, e.g. telephone traffic, radioactive disintegration, airport traffic, estimating the dustiness of the air, etc.

This is perhaps as far as probability can usefully be discussed in a first treatment. The course outlined above should be supplemented by additional numerical questions and by some further experimental work to resolve any doubts expressed by the class.

11. Sampling

Sampling is so common a practice that most schoolchildren will have had some experience of it, though few will have given the subject any serious thought. The following notes outline a suggested first treatment.

Why do we take samples of things ? One reason is to save the time and energy that would be needed to test the whole bulk from which the sample is drawn, e.g. a wholesale dealer buying fruit and vegetables by the ton cannot afford to inspect every apple or potato he buys—he inspects merely a small part of the produce in some, perhaps not very systematic, way. Another reason is that it is sometimes impossible to sample a property of an object without damaging or even destroying the object, e.g. the greengrocer cannot bite each of his apples to see if they are sweet, nor can a manufacturer of electric light bulbs test all his lamps to ensure that they have the guaranteed length of life. Hence sampling is a common and important way of estimating some property of a large bulk, or 'universe', by testing a small fraction of the bulk.

What are we assuming when we take a sample? It is clear that we generally assume that the sample is fair, or 'representative', that its properties are typical of the universe from which it is drawn. This requirement is not easy to ensure. When a sample is not typical or fair in some respect we say that it is 'biassed'. It is one of the tasks of statisticians to show how bias can be avoided and what risks we take by assuming that any sample is a fair one.

Suppose we wished, for some purpose, to select a sample of 20 boys from this school; how could they be selected to give a fair sample? Would it be fair to select them by going to the entrance gate in the morning and counting the first 20 that we see? No, because boys often come to school in groups, from the same bus or train, from the same district, from the same form, and so on; such a selection would be unlikely to give a fair sample. Would it be fair to ask one of the masters to name 20 boys? Probably not; he would tend to name the boys known to him because they learn Greek or play in his cricket team. No, we need a method which we feel to be independent of personal bias and which gives every boy an equal chance of being selected. A well-known method of achieving this aim, used in lotteries, is to put the names (or numbers corresponding to them) of all the people concerned into a hat or drum, mix them well together, and then pick out as many names as are required. If this process is fairly carried out, each person has an equal chance of being selected; such a selection is called a 'random' sample.

There is a serious objection that might be raised here. The random sample could give a result which is far from representative, e.g. it is possible, though improbable, that the random sample might give us 20 prefects or 20 members of one form. We shall return to this point later, but meanwhile this possibility could be avoided if we decide to take one boy from each of 20 forms, selecting both the boys, and perhaps the forms, in a random way. There are many possible variations of this kind and each one has to be judged on its merits.

Let us briefly consider various methods of sampling and see if we can find any sources of bias in them:

(*a*) A sample of milk for test is taken by opening a milk bottle selected at random and filling a test-tube from it.

(*b*) A 'public opinion poll' on the question of the Sunday opening of cinemas is conducted by asking, on a Sunday, the opinions of passers-by in the street near a cinema which is open.

(*c*) A newspaper editor takes a poll on a political matter by discussing it with 100 regular readers of his newspaper by telephone.

(*d*) A man samples the potatoes in his garden by digging up a root at the end of each of three rows.

It is very difficult to avoid bias in any method involving personal selection; we are often unaware that we may be biassed, and even if we do realise it that does not mean that bias is avoided. As the method of picking names or numbers out of a hat is not practicable if the universe to be sampled is very large, we need a similar random method if we are to be sure of eliminating personal bias. Such a method is provided by the use of a prepared table of 'random numbers'. Though tables of random numbers for this purpose have been published, it is not difficult to make a small set for our own use. We could, for example, draw cards one by one from a pack of playing cards, ignoring or discarding the 'court' cards, counting 10's as 0's, and shuffling the pack before each draw. A quicker method is to use a small 10-sided metal spinning top. In either method the probability of any particular digit being selected is 0·1. The set shown in Table 6 was obtained by the first method.

To pick a random sample of 8 boys from a class of 35 each boy is given one of the numbers from 1 to 35 in any convenient way. The table is then consulted. It can be entered at any point and the numbers can be read in any direction. For our present purpose the quickest method would be to read along the top line, mentally reading the numbers in pairs, thus: 37, 45, 09, 34, 36. Note any that are 35 or less, rejecting the others,

until we have eight numbers, e.g. 9, 34, 8, 22, 14, 31, 2 (repeated), 16. In this case the repeated 2 is ignored if we wanted 8 different boys, but as repeated numbers are always possible

TABLE 6

Random numbers

3	7	4	5	0	9	3	4	3	6
0	8	2	2	5	7	9	8	5	3
1	4	8	3	5	8	9	9	3	1
6	5	5	4	9	3	0	2	8	5
8	4	0	2	9	7	4	0	1	6
8	8	9	1	4	5	8	9	6	4
3	3	0	7	9	1	5	6	3	6
5	5	6	1	4	6	6	1	9	5
3	6	2	8	8	3	4	6	7	3
8	6	7	5	1	7	3	3	9	5

the procedure to be adopted if they do occur should be decided beforehand.

It would be disappointing if no boy objected to the use of Table 6 (or its equivalent made in class) on the grounds that it contains too many 3's or too few 2's. The frequencies of the digits occurring in the Table are:—

Digit	0	1	2	3	4	5	6	7	8	9	Total
Frequency . .	6	8	5	15	10	14	12	7	12	11	100

The objection cannot be upheld, at least in this example. The table is in fact our first random sample. The method of making it is analogous to the usual method of drawing names from a hat, except that the selected names are always returned after being noted. The 'hat' therefore never empties however long we continue drawing. There is therefore no reason to expect exactly equal frequencies of digits after any specified number of drawings, though we should expect the ratios of the frequencies to become more and more nearly equal to 1 if the process were continued indefinitely. Most of the remaining

doubts about the unequal frequencies can be dispelled by considering the frequencies of another 100 drawings, e.g. the frequencies produced by the same pack of cards in another set of 100 random numbers were:

Digit . . .	0	1	2	3	4	5	6	7	8	9	Total
Frequency . .	9	5	10	11	16	8	7	11	8	15	100

So much for a technique of random sampling. The next questions are: What can we deduce from a sample? To what extent can we rely on a random sample giving us a fair representation of the bulk? In part, the answers to these questions can be found experimentally. Usually we take a sample because it is difficult or impossible to examine the whole universe from which it is drawn. As an experiment we are going to reverse the process; we are going to make up a universe of a very simple kind and are then going to sample it and see what we get. (The apparatus required is a bag of marbles, or of beans, or of counters, in two colours. There are several varieties of French beans of about the same size and shape but of different colours.) It is assumed that black and white beans are used and that they are contained in a deep bag into which a boy can put both hands.

Here are two heaps of beans, 800 black and 200 white. If the two heaps are put together in the bag and are thoroughly mixed we know that the percentage of white beans is exactly 20. We are now going to see what happens when we take random samples from the bag. (Boys are invited to count out samples of 5, keeping both hands in the bag and counting from one to the other. The sample is brought out and the number of white beans is noted. The sample is then returned to the bag. If the sample results are recorded consecutively they can later be grouped in consecutive pairs, etc., to see the effect of taking larger samples.

The results of 100 (or more) samples are then expressed as a frequency distribution. The expected results, given approximately by the binomial distribution, can be calculated (as a

check only) from the expansion of $100(0{\cdot}8 + 0{\cdot}2)^5$. Expected and observed results are shown in Table 7. Some boys will still be surprised to find that the samples of 5 do not always contain precisely 1 white bean. In fact, the correct proportion of beans was obtained only 46 times in 100, so that for this universe the probability of getting the correct proportion in one sample is about 0·46 or less than $\frac{1}{2}$. A single sample of 5 would therefore

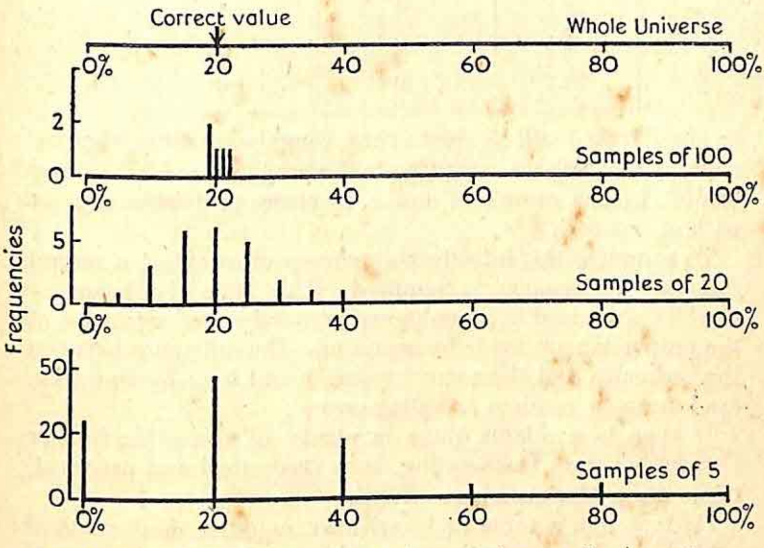

FIG. 9. Effect of increasing sample size on estimates of a proportion.

often mislead us, and mislead us seriously since the probability of finding no white bean at all in a single sample is about $\frac{1}{3}$. A single sample of 5 is therefore not satisfactory for this particular universe. How can we be sure of getting better results? An obvious way is to take larger samples. The larger the sample the closer are the estimates grouped about the correct value as is shown in Fig. 9, which illustrates the effect of taking groups of four, twenty and 100 successive samples of 5.

TABLE 7

Results of sampling experiment

No. of whites	Frequency expected	Frequency observed
0	33	29
1	41	46
2	20	21
3	5	3
4	1	1
5	0	0

In later work it will be shown that, roughly speaking, the precision with which we can estimate the proportion of 'black' and 'white', from a sample of size n, increases proportionally not with n, but with $\sqrt{n}$.

To complete this introductory survey of sampling a second sampling experiment is required. This time the beans or marbles are mixed in an unknown proportion and estimates of the proportion are made by sampling. The difference between the estimates and the correct value, found later by counting, are known as 'random sampling errors'.

If time is available there is plenty of scope for further elementary work in sampling, both theoretical and practical. Some suggestions are here offered:

(1) Examine a table of logarithms (or other mathematical tables) and discuss whether it could be used to provide random numbers.

(2) What is the defect of the following method of producing a table of random numbers: Throw two dice and add their scores; ignore the 12's and subtract 2 from all other scores?

(3) Estimate the mean height (weight, marks, etc.) of a class from a random sample and compare it with the correct result.

(4) Ask each member of a class to write down (a) his own height, (b) his estimate of the mean height of the class. Examine the results to see if tall boys over-estimate, and short boys under-estimate, the mean.

(5) From a large number of picked common daisies take a small sample and estimate the number of ray florets in each flower. Compare the result obtained from the sample with the correct result (i.e. correct for the whole collection).

12. Conclusion

For the teacher who has had no academic training in Statistics, or practical experience of the subject, these notes will need to be supplemented by reference to elementary text-books of statistics. There are several suitable books available but it would be invidious to attempt to give a selection from them. Some will be found on the shelves of any public library or technical bookshop and a brief examination of them should enable the teacher to decide which are appropriate to his particular needs.

SECTION II

TEACHING NOTES ON THE G.C.E. SYLLABUSES IN STATISTICS

1. INTRODUCTION

The notes that follow are intended to help those teachers of mathematics who, with no previous experience of Statistics, would like to begin teaching the subject as a branch of applied mathematics to G.C.E. candidates. The order and selection of the notes have not been determined by any one of the existing syllabuses but broadly cover all of them, so that for any particular examination only some of the topics mentioned will be necessary. The notes must of course be considered merely as a supplement to text-books on the general theory and it is assumed that the reader has some of these books available. Even when text-books written specially for school use begin to appear it will still be impossible to teach the subject satisfactorily without reference to more advanced works. Two works of reference on the elementary theory suitable for school work are:

Yule and Kendall—*An Introduction to the Theory of Statistics.* Griffin. 14th Ed. (34*s.*)

Weatherburn—*A First Course in Mathematical Statistics.* Cambridge. (15*s.*)

These two books at least should find a place in the mathematical library of schools where Statistics is taught as an examination subject. Reference will be made to these books in the notes, using for them the abbreviations YK and W. It is important to have in addition some source of topical data, e.g. copies of the *Annual Abstract of Statistics* (H.M.S.O. 10*s.*) or of the *Monthly Digest of Statistics* (H.M.S.O. 2*s.* 6*d.*).

Some practical work in sampling and in verifying probability distributions is essential if the fundamental ideas of statistical method are to be grasped by schoolchildren. The minimum

apparatus required is very simple—packs of cards, dice, bottles or bags containing prepared populations of beans, marbles or counters, etc.—but there is scope for ingenuity in designing simple pieces of apparatus that illustrate particular ideas or speed up sampling techniques. Any co-operation that the science staff can be persuaded to give is invaluable in providing 'live' statistical data: in return, the school teacher of Statistics must show his science colleagues, and particularly the biologists, how statistical methods can assist the teaching of science.

It is commonly thought by teachers of mathematics that the mathematics of Statistics is in some way very different from the mathematics that is usually taught in Sixth Forms. In fact Statistics provides a field for the application of all branches of Sixth Form pure mathematics except trigonometry and pure geometry and, as a calculus of the discrete variable, Statistics is a corrective to the present over-emphasis on the calculus of the continuous variable. Another criticism is that Statistics is not 'mathematical'; in fact the content of any school theoretical Statistics syllabus is limited by the content of the parallel pure mathematics course. To extend the theory covered by these notes some acquaintance with Beta and Gamma functions, double integration, hypergeometric series, etc., is required. One aim of these notes is to indicate how closely elementary Statistics is related to the usual school course in pure mathematics.

2. Descriptive Statistics

2(a) *Tabulation and Graphical Representation*

It is unnecessary to attempt a systematic and exhaustive treatment of these topics to begin with. Though they are important, tabulation and graphical representation are necessary concomitants of all subsequent work, and therefore opportunities of discussing them arise throughout the course. A high standard of neatness and lay-out of tables must be demanded, primarily as a means of reducing arithmetical errors. Paper ruled in $\frac{1}{4}$-inch squares is useful; a foolscap loose-leaf file is convenient for holding exercises. Practice in the tabulation of

raw unclassified data is required, but opportunities arise in later work, e.g. sampling experiments. Some of the difficulties of tabulating and classifying data are illustrated by the foot-notes that accompany the tables in the *Annual Abstract* and *Monthly Digest*.

2(*b*) *Frequency Distributions and Histograms*

(i) The words *variable* and *variate* are not used consistently in Statistics; some writers use them as synonyms, others define a variate as a variable which has a probability distribution. The distinction is not important in school work and if it is made the expression 'bivariate distribution' might imply more than is intended. It is therefore best to use both words and to use them as synonyms.

(ii) Some practice in the selection of suitable class limits should be left to the pupils; for school work a total of 20 classes is enough. It is conventional to add $\frac{1}{2}$ to the frequencies of adjacent class frequencies should the value of the variable coincide exactly with the value dividing the two classes.

(iii) The fact that grouping is a form of approximation is sometimes lost sight of. It is a useful exercise to begin with a large number of accurate readings of a continuous variable (e.g. barometric height) and to reduce them from, say, 200–300 classes in steps finally to 2 or 3. The increasing roughness of the approximations is demonstrated by the corresponding histograms.

(iv) The use of the descriptive terms *symmetrical*, J*-shaped* U*-shaped*, *positive* and *negative skew*, *unimodal* (or *one-humped*), *bimodal* (or *two-humped*) should be encouraged.

(v) The fundamental property of the histogram is that the *area* of any column is proportional to the corresponding frequency. This point may be overlooked if frequency distributions with equal class intervals only are used; the *Annual Abstract* provides many examples of distributions with unequal class intervals.

(vi) *The frequency polygon* deserves little more than a mention.

(vii) At this stage the drawing of two or three histograms is sufficient to illustrate all the essential points and to give pupils sufficient confidence in their use and interpretation. They will be used continuously in later work.

(viii) Logarithmic scales and logarithmic graph paper are now so widely used that they should be introduced when a suitable opportunity arises. Some manufacturers offer sample packages which include many types of logarithmic graph papers that are useful in other mathematical work (e.g. for home-made slide rules).

2(c) *Averages; Mode, Median, Mean* (*YK, Ch.* 5)

(i) In comparing two or more frequency distributions, or their histograms, some method is required of stating their *location* along the line which represents the values of the variate. It is useful to remember that we usually compare distributions whose histograms are roughly similar in shape.

(ii) The *mode* is strictly applicable only to frequency curves, so that at this stage it is better to use the *modal group*; its limitations are obvious.

(iii) The *median* should be found both from its definition as the middle value (including the case of distributions with even numbers of values) and by graphical interpolation from the *cumulative frequency diagram*. In drawing this diagram note that the points corresponding to successive cumulative frequencies are plotted on the ordinates corresponding to the *upper* limits of the class intervals. The points are joined by a smooth curve (the so-called 'ogive').

(iv) The general formula for the *mean* of a frequency distribution is an application of the *weighted mean*. Unless the use of the symbol Σ is already familiar some introductory exercises on its use are required.

(v) The lay-out of the first computations of the arithmetic mean of a frequency distribution is described in Section I, p. 24.

(vi) The *geometric mean* is worthy of mention if only to emphasise the arbitrariness of our choice of averages, and because

it may be needed for Index Numbers. Note that the logarithm of the geometric mean of a number of variables is the arithmetic mean of the logarithms of the variables.

(vii) With the possible exceptions of some model examples to illustrate computing techniques, all exercises should be planned to provide some discussion. Averages are used as a means of comparison; the average of a single distribution unrelated to any other information is of little value or interest.

(viii) Class exercises selected with a view to further analysis will save the tedium of unnecessary repetition of computations if they are filed and kept.

(ix) The check provided by the identity $\Sigma f(d+1) = \Sigma fd + \Sigma f$ should be demanded as a matter of routine.

(x) Some of the algebraic properties of the arithmetic mean should be noted. It is important to translate the symbols into words whenever they are used. Thus, if m is the mean of n values of x, then by definition $m = \Sigma x/n$. The identities $\Sigma (x - m) = 0$, 'the algebraic sum of the deviations about the mean is zero', and $\Sigma x = nm$, 'the sum of n variables is equal to n times their mean value', are frequently used in later work and must be recognised at once.

(xi) For pupils who can appreciate it, the analogy between the formula $m = \Sigma fx/\Sigma f$ for the mean of a distribution, and $\bar{x} = \int yx\,dx/\int y\,dx$ for the abscissa of the mean centre of the plane area enclosed between the curve $y = f(x)$ and the x-axis, should be pointed out. Otherwise, the determination of the arithmetic mean of a frequency distribution by finding the ordinate about which its histogram, cut out of cardboard, balances on a ruler edge, emphasises 'what the arithmetic mean is doing'. Pupils with a knowledge of mechanics will understand why the expression $\Sigma f(x - m)/n$ is called the 'first moment of the distribution about the value m'.

2(*d*) *Moving Averages* (*Section* 1; *YK*, *Ch.* 26)

(i) In a school course the analysis of time series will usually be limited to the use of the *unweighted moving average* as a

means of 'smoothing' time series, i.e. of separating random fluctuations from the long-term trend and from seasonal and other periodic changes. The examples should be selected from data of which the background is known so that interpretation of the results is possible. It is a subject which has many interesting facets for mathematicians and the temptation to explore it in class time may have to be resisted.

(ii) If n is the 'span' of an unweighted moving average the $(r + 1)$th mean is obtained from the rth mean by the use of the formula $m_{r+1} = m_r + \frac{1}{n}(x_{r+n} - x_r)$. It pays to tabulate the work by setting out $x_1, x_2, x_3, \ldots$ in rows of n terms, putting the second row above the first, and so on. The differences $(x_{r+n} - x_r)$ can then be written down in sequence with the appropriate sign. As a check the final average should be computed in the usual way.

2(*e*) *Weighted Means* (*Section* 1; *YK*, *p.* 332)

Practical exercises on weighted means are afforded by traffic counts on two roads, on the same road at different times of the day, or on different days. Comparison of the crude counts may be misleading if the proportions of heavy and private vehicles are markedly different. The devising of a suitable system of weighting different types of vehicle is a useful exercise. Information about the numbers of licences issued and the classification of vehicles is given in the *Annual Abstract*.

2(*f*) *Index Numbers* (*Section* 1; *YK*, *Ch.* 25)

For the mathematical statisticians it is sufficient to limit the study of Index Numbers to the fact that they are weighted arithmetic or geometric means of price relations, using topical examples. The economists will attach more importance to this topic and a more detailed treatment may be required: that provided by YK includes an account of time-reversal and circular tests and should meet all school needs.

2(*g*) *Vital Statistics* (*YK*, *pp*. 333–337)

Another topic of social interest that depends on the application of the weighted mean is the correction of vital statistics. For example, some health resorts on the south coast of England may be found to have a higher crude death rate (measured as deaths per 1000 of the population) than some of the smoky towns in industrial areas. One reason is that the population distributions are different; people in the industrial towns are comparatively young workers, whereas the coast resorts attract many elderly retired people. Briefly, the standardised death rate is obtained by weighting the death rates in each age class by weights based on the age distribution of a standard population. Similar methods are applied to birth rates.

2(*h*) *Measures of Dispersion* (*YK*, *Ch*. 6)

(i) A measure of the *spread* or *dispersion* of a distribution is the second characteristic by which we compare one frequency distribution with another.

(ii) The *range* is usually mentioned only to be dismissed, though it is used effectively in Quality Control of mass production processes (Section I. 8). It is not generally realised that the range becomes less efficient as the sample size increases.

(iii) The *semi-interquartile* range is usually determined graphically from the cumulative frequency curve. It provides an opportunity of mentioning *percentiles* and their uses, e.g. with examination marks, anthropometric data, etc., but it is of little mathematical interest.

(iv) The *mean deviation* is the mean of the absolute deviation. Though it is a minimum when the deviations are measured from the median it is usually more convenient to measure them from the mean. For a first treatment it is usually good enough to measure the deviations from the centre of the class-interval which contains the mean, though the correction is not difficult.

(v) The *standard deviation*, *s*, is defined by $s^2 = \Sigma fd^2/n$ where d is the deviation from the mean. It should be pointed

out that it is a particular case of the more general *root mean-square deviation*, that it is in fact the minimum value of that quantity. This last relation should be proved before any computation is attempted, as it is required to explain the correction to be made in the computation when it is inconvenient to measure the deviations exactly from the mean. If the pupils have already met the Principle of Parallel Axes in Moments of Inertia it is helpful to point out the analogy between the s.d. and the radius of gyration of a plane lamina of the same shape as the histogram about a line in the plane which is parallel to the y-axis and which passes through the mean centre of the histogram.

(vi) The tabulated computation of the s.d. is described in most elementary text-books. Note that it is usually quicker to find fd^2 from $fd \times d$ than from $f \times d^2$. The check based on the identity $\Sigma f(d+1)^2 = \Sigma fd^2 + 2\,\Sigma fd + \Sigma f$ should be demanded. As the computation is tedious, examples set as class-work should be modest unless the work can be shared out.

(vii) Sheppard's correction for grouping is not required (nor is it *generally* applicable) but it is important to point out that grouping does introduce small errors.

(viii) It is of little interest to compute any of the various measures of dispersion (except as examples of computing techniques) unless they are used in comparing distributions.

(ix) A comparison of the usefulness of the several measures of dispersion is best deferred until the pupils have had some experience of all of them.

(x) Note that the dimensions of the measures of dispersion are those of the variable.

2(j) *Standard Measure*

(i) Distributions of marks, particularly of intelligence-test scores, are sometimes compared by converting the 'raw' scores to 'standard' scores, which are obtained by expressing the deviations of the raw scores from their mean as multiples of the s.d. Sometimes a further conversion of the marks to a standard scale with a given mean and s.d. is performed. One

or two numerical exercises, as extensions of the computation of the s.d. of a distribution of marks, should suffice, e.g. Find the mean and s.d. of a given distribution of marks and convert them (or some of them) to a distribution which has the mean 100 and s.d. 15.

(ii) The *coefficient of variation*, $100s/m$, is sometimes used to compare variabilities, e.g. whether boys of 14 years vary more in height or in weight than boys of 17 years. It is obvious that it can be used only when the ratio s/m is small.

2(*k*) *Variance*

(i) The *variance* of x, $V(x)$, is the square of the standard deviation. It is a *statistic* of great importance in more advanced work. Some algebraic exercises on it provide results that can be applied to elementary work.

(ii) In the combining of two sets of n_1 and n_2 observations of a variable with means m_1, m_2, and variances V_1 and V_2, the variance V of the combined sets of observations can readily be shown to be given by

$$(n_1 + n_2)V = n_1V_1 + n_2V_2 + \frac{n_1n_2}{n_1 + n_2}(m_1 - m_2)^2$$

(Cf. the combined M.I. of two masses M_1, M_2, with moments of inertia I_1, I_2, about their mean centre.)

If $n_1 = n_2$ and $m_1 = m_2$, then

$$V = \tfrac{1}{2}(V_1 + V_2)$$

(iii) If $z_r = ax_r + by_r$ where a and b are constants, then

$$\Sigma\, z_r = a\,\Sigma\, x_r + b\,\Sigma\, y_r$$

and therefore $\bar{z} = a\bar{x} + b\bar{y}$, where $\bar{x}$, $\bar{y}$, $\bar{z}$, are the means of the x's, y's and z's. Hence

$$\Sigma\,(z_r - \bar{z})^2 = \Sigma\,\{a(x_r - \bar{x}) + b(y_r - \bar{y})\}^2$$
$$= a^2\,\Sigma\,(x_r - \bar{x})^2 + b^2(y_r - \bar{y})^2 + 2ab\,\Sigma\,(x_r - \bar{x})(y_r - \bar{y})$$

If x and y are independent, the product term may be neglected in comparison with the others, and so

$$V(z) = a^2V(x) + b^2V(y)$$

In particular if $a = b = 1$, or if $a = -b = 1$, then

$$V(z) = V(x) + V(y)$$

(iv) The results: (*a*) That $\Sigma\,(x_r - \bar{x})(y - \bar{y})$ can be neglected in comparison with $\Sigma\,(x_r - \bar{x})^2$ or $\Sigma\,(y_r - \bar{y})^2$; (*b*) That $V(z) = V(x) + V(y)$ for the sum or difference of x and y, where x and y are independent variables, needs practical verification. Pairs of the digits 1, 2, . . . 9 drawn from two card packs or from a table of random numbers are suitable; if the zeros are ignored the computation becomes simpler since $\bar{x} = \bar{y} = 5$ and fractions are avoided.

2(*l*) *Continuous Distributions* (*W*, *p.* 12)

(i) It is helpful to mathematical pupils familiar with the notations of the calculus if the descriptive statistics of continuous frequency distributions are given in parallel with those of discrete distributions.

(ii) The limit of a histogram in which the class intervals become infinitesimally small as the frequencies tend to infinity is a smooth curve which we assume can be expressed in the form $y = f(x)$. If the continuous variable x can take all values within the range a to b and if $\int_a^b f(x)\,dx = 1$, so that the area under the curve in this range is unity, the curve $y = f(x)$ is called the *relative frequency curve* of the distribution. The function $f(x)$ is called the *relative frequency density*. The area property of the histogram is retained, i.e. the relative frequency of the interval h to k is $\int_h^k f(x)\,dx$.

(iii) For the relative frequency distribution $y = f(x)$ we have the mean,

$$\mu = \int_a^b x\,f(x)\,dx$$

the variance,

$$\sigma^2 = \int_a^b (x-\mu)^2 f(x)\,dx$$
$$= \int_a^b x^2 f(x)\,dx - \mu^2$$

The cumulative frequency curve is given by

$$y = \int_a^x f(x)\,dx$$

and so the median is given by

$$\int_a^x f(x)\,dx = \tfrac{1}{2}$$

and so on.

(iv) The functions $3x^2\,(0 \leqslant x \leqslant 1)$, $e^{-x}\,(0 \leqslant x < \infty)$, $\frac{2}{\pi}\sin^2 x (0 \leqslant x \leqslant \pi)$, $\frac{1}{\pi}x \sin x (0 \leqslant x \leqslant \pi)$, $\frac{1}{\pi}\cdot\frac{1}{1+x^2}\,(-\infty < x < \infty)$ provide values of $f(x)$ suitable for simple exercises.

3. Probability and Sampling

3(*a*) *The Laws of Probability*

(i) Some revision of permutations and combinations may be needed.

(ii) One mathematical definition of probability is: If an action can entail any one of n equally likely results, and if m of these results entail the occurrence of an event E and the remainder do not, then the ratio m/n is the probability of E.

An empirical definition of probability is: If an event E, one of the mutually exclusive results of a trial, is found to occur m times in n trials, then the limit of the ratio m/n as n tends to infinity is the probability of E.

There are logical objections to both definitions though schoolboys are unlikely to raise them. A criticism of the first definition is that 'likely' is a synonym for 'probable' and that therefore the definition is circular; in the second definition it is assumed that the limit of m/n exists in the mathematical

sense. In a first school course of Statistics both definitions are required and are used as equivalents. A critical examination of them must await experience of their application by the pupils. For this later review the short discussion in Aitken, *Statistical Mathematics*, pp. 5–15, is useful, and Bernouilli's theorem as presented in W, pp. 33, 34, is of interest.

(iii) Proofs of the theorems of *total* ('either-or') and *compound* ('both-and') probability are given by W (pp. 22, 23). These theorems are sometimes known as the 'addition and product rules'. A third rule—the 'at least one' rule—is worth giving: If the probabilities of n independent events are p_1, p_2, . . . p, then the probability of at least one of the events occurring is $1 - (1 - p_1)(1 - p_2) \ldots (1 - p_n)$, e.g. the chance of at least one 6 in the tossing of n unbiassed dice is $1 - (\frac{5}{6})^n$.

(iv) The useful concept of the *expected value* or *expectation* for the mean value of a variable should be introduced (W, p. 24), e.g. the expectation in a throw of two dice is 7.

3(*b*) *Some Approximations*

A list of approximations that are used in subsequent work is appended; some numerical examples of their accuracy are needed to make their validity convincing.

(i) Logarithmic approximations.

$$\log_e (1 \pm x) \approx \pm x - \tfrac{1}{2}x^2$$

for small values of x is well known.

(ii) Exponential approximations.

(*a*) $e^x \approx 1 + x$, $e^{-x} \approx 1 - x$ if x is small.

(*b*) When x is very small, $(1 + x)^n \approx e^{nx}$, $(1 - x)^n \approx e^{-nx}$.

(*c*) $e \approx (1 + n)^{n+\frac{1}{2}}$. The error is of the order of $1/12n^2$ for any value of n; e.g. when $n = 2$ the approximation gives $e \approx 2{\cdot}7559$.

(iii) Stirling's approximation.

$$n! \approx (2\pi n)^{\frac{1}{2}} e^{-n} . n^n$$

for large integral values of n. There are various proofs but all

are either very long or beyond the usual school course. It is used mainly to simplify ratios of factorials; in the form

$$\frac{n!}{m!} \approx \frac{e^{-n}n^{n+\frac{1}{2}}}{e^{-m}m^{m+\frac{1}{2}}}$$

it can be verified by the use of the approximation (*c*).

3(*c*) *The Binomial Distribution*

(i) Establish the binomial distribution by considering a repeated trial generally (W, p. 28).

(ii) There are various proofs of the formulae $\mu = mp$, $\sigma = (npq)^{\frac{1}{2}}$ (W, p. 28, 46; YK, p. 174). More direct proofs are simplified if the results below are first established:

$$(a + x)^n = \Sigma_0^n \binom{n}{r} a^{n-r}x^r$$

Differentiate with respect to x,

$$n(a + x)^{n-1} = \Sigma_1^n r\binom{n}{r} a^{n-r}x^{r-1} \qquad . \quad . \quad (1)$$

Multiply both sides by x and differentiate again. On reducing the left-hand side,

$$n(a + x)^{n-1}(a + nx) = \Sigma_1^n r^2\binom{n}{r} a^{n-r}x^{r-1} \qquad . \quad (2)$$

(iii) It is important to show that the binomial distribution for unequal values of p and q tends to a symmetrical form as the exponent increases (YK, p. 172).

(iv) Numerical examples on the binomial distribution will be concerned to establish its use as a probability distribution by computing probable frequencies in simple cases and comparing them with observed and experimental frequencies.

(v) Show that for large values of n (100 or more) about 99% of the area under the binomial histogram is included between the ordinates at $\mu \pm 3\sigma$, and about 95% between the ordinates at $\mu \pm 2\sigma$.

3(*d*) *The Poisson Distribution*

(i) The derivation of the Poisson distribution from the bi-

nomial usually requires the use of Stirling's approximation for $n!$ (YK, p. 190; W, p. 64). Alternative methods:

(*a*) It is required to show that, when $n \to \infty$ and $np = a$, a constant, the limiting value of $\binom{n}{r}p^r q^{n-r}$ is $e^{-a}a^r/r!$.

$$\begin{aligned}\binom{n}{r}p^r q^{n-r} &= \binom{n}{r}p^r(1-p)^{n-r} \\ &= \frac{n!}{(n-r)!r!}\left(\frac{a}{n}\right)^{-r}\left(1-\frac{a}{n}\right)^{n-r} \\ &= \frac{n!}{(n-r)!n^r}\left(1-\frac{a}{n}\right)^{-r}\left(1-\frac{a}{n}\right)^{n}\frac{a^r}{r!} \\ &= \frac{n!}{(n-r)!(n-a)^r}e^{-a}\frac{a^r}{r!}\end{aligned}$$

and the limit of the coefficient for finite values of a and r can be shown directly to be 1.

(*b*) $$q^n = (1-p)^n = \left(1-\frac{a}{n}\right)^n \approx e^{-a}$$

$$nq^{n-1}p = np(1-p)^{n-1} = \frac{a}{q}\left(1-\frac{a}{n}\right)^n \approx ae^{-a}$$

$$\frac{n(n-1)}{2!}q^{n-2}p^2 = \frac{n(n-1)}{2!}\cdot\frac{p^2}{q^2}\left(1-\frac{a}{n}\right)^n \approx \frac{a^2}{2!}e^{-a}$$

Similarly the $(n+1)$th term is $\frac{a^n}{n!}e^{-a}$ for sufficiently large n.

Discussion of the error term need not delay some illustrative examples.

(ii) The formulae $\mu = a, \sigma = \sqrt{a}$, can be derived directly from the known binomial results $\mu = np$, $\sigma = (npq)^{\frac{1}{2}}$, or by taking first and second moments about the origin (YK, p. 191). Note that the variance is numerically equal to the mean or expectation.

(iii) Histograms of some Poisson distributions should be

drawn, e.g. $a = \frac{1}{2}, 1, 3, 10$, noting the trend towards symmetry as a increases.

(iv) In the absence of tables of the Poisson distribution numerical examples are limited to the comparison of observed and computed frequencies of distributions in which a is not greater than about 3, and to simple examples of the type: It is known that, on average, $2\frac{1}{2}\%$ of a mass-produced article are defective: what is the probability of finding 3 or more defective articles in a package of 80?

[Expected number per package $= 2 = a$. Hence

$$P = 1 - e^{-2}\left(1 + \frac{2}{1!} + \frac{2^2}{2!}\right) = 1 - 5e^{-2}$$
$$= 0{\cdot}3245]$$

(v) The Poisson distribution applies to the occurrence of comparatively rare random events, and illustrative data sometimes take a long time to collect. Some practical exercises that fairly quickly give results (at least 100 samples are required) are suggested below. The unit time intervals, etc., depend on local conditions but should be arranged to give an expectation of not more than 3 or 4 occurrences per unit time interval, etc.

(*a*) See games with car registration numbers in Section I, p. 41.

(*b*) A few grains of sand are spread over graph paper with $\frac{1}{4}$-inch squares. The number of grains in each square is noted. (Cf. blood-counts and estimation of dust particles in the air.)

(*c*) Note the frequency of occurrence of number-pairs or triplets in equal sections of a table of random numbers.

(*d*) Samples are drawn from a prepared bag of beans, marbles, etc., using a measure instead of counting.

(*e*) The frequency of occurrence of telephone calls or of visitors in suitable time intervals may sometimes be made available with the co-operation of the office-boy.

3(*e*) *Simple Sampling of Attributes* (*Section I; YK, Ch.* 16, 17, 23)

(i) The first need is to establish meanings for the terms

random, *bias* and *simple sampling*; many useful ideas are given in the YK reference quoted. Secondly, a clear distinction should be drawn between the sampling of *attributes* and the sampling of *variables*.

In simple sampling it is assumed that the selection is random and, further, that the chance of a success, p, is constant for each item drawn. Each item must therefore be replaced before the next item is selected, or the population from which the sample is drawn must be very large, e.g. the probabilities of hearts in 4-fold samples from a pack of cards are:

No of hearts in sample	0	1	2	3	4
With replacement . .	0·3164	0·4219	0·2109	0·0469	0·0039
Without replacement .	0·3038	0·4388	0·2135	0·0412	0.0026

(Hogben, *Chance and Choice*, Vol. 1, p. 97.)

(ii) Since in simple sampling of attributes we are concerned with repeated independent trials, each with a constant probability of success, the probability of r successes in an n-fold example is given by the $(r + 1)$th term of the expression of $(p + q)^n$. The binomial distribution then becomes the sampling distribution, i.e. the standard or model against which our sampling of attributes is to be compared. We already know that for the binomial distribution $\mu = np$ and $\sigma = (npq)^{\frac{1}{2}}$, or if we use proportions instead of actual frequencies, $\mu = p$ and $\sigma = (pq/n)^{\frac{1}{2}}$.

(iii) In a first course there are two types of question to which the theory of the simple sampling of attributes gives an answer, though only in terms of probability:

(*a*) If an n-fold sample gives a proportion, p, of items with a certain attribute, what is the proportion of items with that attribute in the sampled population?

(*b*) If in a trial experiment to test a hypothesis a proportion p_1 of items with a certain attribute is found in an n-fold sample, does this result invalidate the hypothesis that the proportion is p?

(iv) In questions of the first type the correct value of p is unknown. The estimate of p provided by the sample is

accepted and is used to determine the standard error of the sampling distribution, i.e. $(pq/n)^{\frac{1}{2}}$. From our knowledge of the binomial distribution for values of n of 100 or more we can then say that the required proportion lies within the range $p \pm 3 \times$ s.e. with a probability of about 0·99.

(v) In questions of the second type the s.e. is computed from p. Then if $|p - p_1| < 3 \times$ s.e., we say that the difference could arise by fluctuation of random sampling or that the difference is not significant at the 3σ level, and that the result affords no evidence on which to reject the hypothesis. If $|p - p_1| > 3 \times$ s.e., we say that it is very improbable that this difference could arise from a random sampling, or that the difference is significant at the 3σ level, and that the difference is sufficiently great to lead us to doubt the hypothesis or to re-examine the conditions of sampling.

3(*f*) *The Normal Distribution*

(i) Though the derivation of the equation of the Normal curve from the binomial distribution is not usually required for examination purposes it can be demonstrated (YK, p. 177; W, p. 64).

(ii) The next step is to investigate the properties of the curve $y = \exp(-x^2/2a^2)$ as an exercise in curve tracing.

(iii) If continuous distributions have not been discussed already a digression will be needed to introduce the terms relative frequency, etc. The area under a probability curve is by convention unity, and therefore the equation

$$\int_{-\infty}^{\infty} k.\exp(-x^2/2a^2)\,dx = 1$$

has to be solved for the constant k. A solution avoiding Gamma functions is given by W, p. 65.

(iv) As the curve is introduced as an approximation to the binomial histogram a numerical and graphical example to illustrate the goodness of fit will be needed.

(v) Teachers will find the use of Arithmetic Probability

graph paper very useful for the preparation or checking of examples on the Normal distribution. The paper is ruled so that the cumulative frequency curve $y = \int_{-\infty}^{x} \exp(-x^2)\,dx$ becomes a straight line. Values of μ and σ for a given distribution can be obtained directly from the graph, or alternatively the paper provides an easy method of preparing a Normal frequency distribution with a given μ and σ.

(vi) For future work, tables of ordinates of, and of areas under, the Normal curve will be required. For most school purposes carefully drawn graphs can give the necessary degree of accuracy and have the advantage over tables of showing more clearly what is being done when reference is made to them. For the ordinates it is necessary to plot values of $\sigma y = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2\sigma^2)$ for a few values of x/σ (e.g. 0, $\frac{1}{2}$, 1, . . . 3, $3\frac{1}{2}$). For the areas it is necessary to evaluate

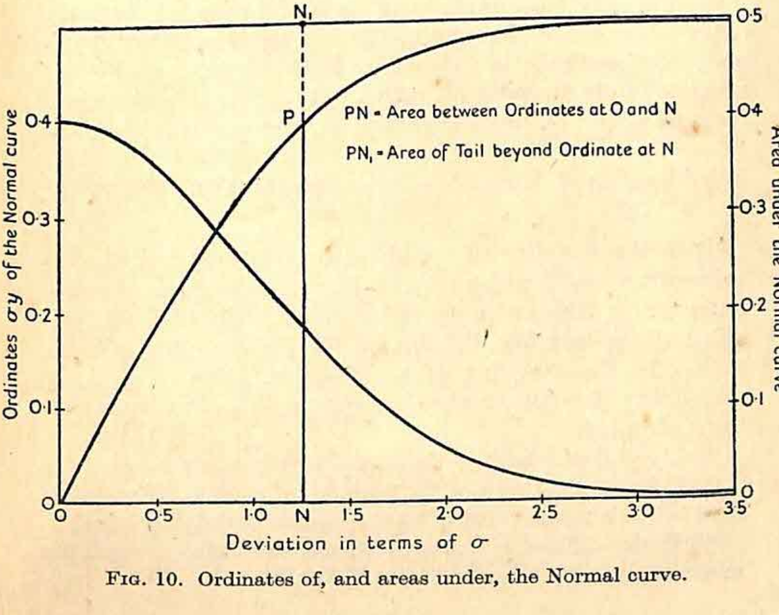

FIG. 10. Ordinates of, and areas under, the Normal curve.

$\int_0^x \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp(-x^2/2\sigma^2)dx$ for the same values of x/σ by expanding the exponential in series and integrating term by term (*vide* Whittaker and Robinson, *The Calculus of Observations*, pp. 180, 181). The form of the graphs shown in Fig. 10 is convenient for reference.

(vii) The Normal is the only continuous distribution listed in most of the syllabuses, but pupils should not be allowed to conclude that it is the only one and that all sampling distributions are Normal. The name 'Normal' can mislead, particularly if it is written 'normal'; a distribution that is 'not normal' can only be 'abnormal'! The name 'Gaussian' is an alternative though it does less than justice to De Moivre and Laplace.

3(*g*) *The Rectangular Distribution*

The rectangular distribution $y = 1/(b - a)$, $a \leqslant x \leqslant b$ can be both continuous and discrete. It is of no special statistical importance, but it is useful to the teacher because its simple properties are easy to investigate both mathematically and practically; its simplicity enables it to be used in the classroom as a kind of 'proving ground' for many statistical ideas.

3(*h*) *Sampling of Variables—Large Samples* (*YK*, *Ch.* 18; *W*, *p.* 116)

(i) In the sampling of variables the items selected have a value of a variable which may be discrete or continuous over a range which may or may not be known. The binomial provided the probability distribution for the sampling of attributes; for the sampling of variables we have to seek the probability distribution that is appropriate to the population being sampled.

(ii) Typical problems in the sampling of variables are:

(*a*) What is the mean of the population sampled if an n-fold sample has a mean value of m and a standard deviation s?

(*b*) Is the difference between an expected mean μ and the mean m given by an n-fold sample significant?

(c) The means of two samples are m_1 and m_2; is the difference of these means consistent with the hypothesis that they were drawn from the same population?

(iii) These questions again can be answered only in terms of probability and to answer them we need to know the *standard error of the mean*.

(iv) A number of n-fold samples drawn from the same population gives a distribution of values of m. This fact should be established practically (for small samples) by, for instance, finding the means of samples of 4, 9, . . . random numbers. Is it possible to determine the characteristics of these distributions of m?

We already know that if $z_r = x_r + y_r$, where x_r and y_r are independent, then $V(z) = V(x) + V(y)$. If m is the mean of $x_1, x_2, \ldots x_r \ldots x_n$, we know that

$$nm = x_1 + x_2 + \ldots + x_r + \ldots + x_n = \Sigma\, x_r$$

and as the values of x are independent, we have by extension of our variance formula,

$$V(nm) = \Sigma\, V(x_r)$$

But $V(x_r) = \sigma^2$, the variance of the distribution of x's from which the sample is drawn.

Therefore $\qquad V(nm) = n\sigma^2$

and $\qquad \text{s.d.}(nm) = \sigma\sqrt{n}$

or $\qquad \text{s.d}(m) = \sigma/\sqrt{n}$

The standard deviation of any statistic of a sampling distribution is called the standard error of that statistic. Hence we have s.e. $(m) = \sigma/\sqrt{n}$, where σ is the s.d. of the universe from which the sample is drawn and n is the sample size. Note that this formula has been obtained without specifying the form of the sampled distribution.

(v) This result gives only the s.d. of the distribution of means, though we have implicitly assumed that the mean of the distribution is the mean of the sample means. We know nothing however of the form of the distribution. It can be shown theoretically that if the sampled population is Normal,

then the distribution of means is also Normal, but the distribution of means of samples drawn from other one-humped distributions at least approximates to the Normal. Though these facts cannot be established theoretically in a school course they can be shown by practical examples to have some basis.

In applying the s.e. of the mean in answering the questions of para. (ii) we can therefore use the 3σ level of significance. If the s.d. of the population is unknown, then the information provided by the samples is pooled; since they are large samples the estimates of σ they provide are not greatly in error.

3(*j*) *Population Parameters and Sample Statistics*

It is conventional to distinguish between parameters which refer to a universe or population, and estimates of them derived from a sample, by using Greek letters for the parameters and corresponding Latin letters for the estimates. Thus m is used to denote an estimate of μ, s to denote an estimate of σ, r to denote an estimate of ρ, and so on. In a school course it is helpful to bear this useful convention in mind from the beginning of the course so that the distinction that has to be made when sampling is considered is not blurred by earlier violations of the convention.

Unfortunately this useful convention breaks down when applied to the usual symbols of the binomial distribution. The letters p and q are conventionally used to denote both the observed and the theoretical proportions of successes and failures; their Greek analogues π and χ could be confusing since they have other established meanings which may be simultaneously required. Other possibilities are to use $p = \text{est}(P)$, $q = \text{est}(Q)$, or $a = \text{est}(\alpha)$, $b = \text{est}(\beta)$, but of course any such convention must first gain general acceptance.

3(*k*) *Levels of Significance and Confidence Limits*

(i) If a sampling distribution is known to be Normal it is possible to be more precise about levels of significance. First it should be noted that levels of significance are purely arbitrary;

the choice depends in part on the nature of the problem, on the importance of any action that depends on the result, on the caution of the investigator. There are two levels which are commonly used, the 5% and the 1% levels. These correspond roughly to the 2σ and 3σ levels of the Binomial distribution, since for the Normal distribution 5% of the area under the curve lies outside the ordinates at $\mu \pm 1{\cdot}96\sigma$, and 1% outside those at $\mu \pm 2{\cdot}58\sigma$. The ordinates at $\mu \pm 3\sigma$ exclude all but 0·27% of the area under the Normal curve. For some purposes levels of 0·1% may be required. Since the level is a matter of choice it is important always to complete any statement about significance by specifying the level used.

(ii) If the sampling distribution is known, but is not Normal, the deviations corresponding to 5% and 1% levels of significance can be determined. If nothing is known of the sampling distribution then at worst the probability that a sample value will differ from the mean by more than $\lambda\sigma$ is not greater than $1/\lambda^2$ (Tchebychef's theorem; W, p. 33). Usually however we can do much better than this; most elementary statistics computed from random samples of non-Normal distributions are approximately Normal if the sample is large.

(iii) Sometimes *confidence* or *fiducial* limits are used. The terms are not strictly synonymous but the distinction is not important in a school course. For example, if for a sample drawn from a normal population $m = 3{\cdot}60$ with a standard error of 0·12, the 95% confidence limits for are $3{\cdot}60 \pm 1{\cdot}96 \times 0{\cdot}12$, i.e. 3·84 and 3·36.

(iv) The form of a statistical question involving significance is sometimes important. The questions: 'Is this coin biassed?', 'Does this coin give more heads than we expect of an unbiassed coin?' are statistically different. If a trial of 100 tosses produced only 10 heads we would have little hesitation in answering 'Yes' to the first question and 'No' to the second. To the first question we apply a 'two-tail' test, i.e. we allow for the area under the probability curve outside both significance limits; to the second question we apply a 'one-tail' test, i.e. we are concerned with the area under only one tail.

3(l) *Sampling of Variables—Small Samples* (*YK, Ch.* 21; *W, Ch. X*)

(i) In considering small samples (i.e. $n < 100$) it is necessary to subject the theory of large sampling to a critical examination for possible errors and then to allow for them. The results are:

(a) First the form of the sampled population has to be specified; the notes that follow apply only to a sampled population that is Normal or approximately so.

(b) The distribution of the means of an n-fold sample drawn from a Normal population (mean μ, variance σ^2) is also Normal (mean μ, variance σ^2/n) whether the sample is large or small. Hence $m = \text{est}(\mu)$ as for large samples (W, p. 119).

(c) The variance of the sample, s^2, gives an estimate that must be corrected for the systematic error that arises from the fact that in computing s^2 the deviations are measured from m and not from μ (which is usually unknown). By making the adjustment it is found that the 'unbiassed' estimate of σ^2 is given by $\Sigma\,(x - m)^2/(n - 1)$ rather than by $\Sigma\,(x - m)^2/n$ (W, p. 131). Note that 'unbiassed' is here given a technical meaning (YK, p. 547).

(d) In the theory of large samples we assumed that the distribution of $\dfrac{(m - \mu)}{s/\sqrt{n}}$ was Normal, or approximately so. This distribution, however, deviates appreciably from the Normal if n is small. The exact distribution of this statistic, called the t-distribution, was first determined by 'Student' and tabulated; the derivation is not required for a school course but is given in W, p. 186.

(e) Using tables of the t-distribution it is possible to determine the significance at any given level of a difference between the mean of a sample and its hypothetical value, or between the means of two small samples, by methods analogous to those used for large samples.

(f) This more exact test applies equally well to large samples as to small ones, but as the sample size increases the t-distribu-

tion approaches the Normal distribution more and more closely. For samples of 100 or more, therefore, the more convenient Normal probabilities are usually applied.

TABLE 8

An approximately normal distribution

($n = 1000$, $\mu = 50$, $\sigma = 10$)

Pop. No.	Value	Frequency	Value	Pop. No.
1	18	1	82	1000
2	20	1	80	999
3	21	1	79	998
4	22	1	78	997
5	23	1	77	996
6	24	1	76	995
7– 8	25	2	75	993–994
9– 10	26	2	74	991–992
11– 13	27	3	73	988–990
14– 17	28	4	72	984–987
18– 21	29	4	71	980–983
22– 26	30	5	70	975–979
27– 33	31	7	69	968–974
34– 41	32	8	68	960–967
42– 50	33	9	67	951–959
51– 61	34	11	66	940–950
62– 74	35	13	65	927–939
75– 89	36	15	64	912–926
90–106	37	17	63	895–911
107–125	38	19	62	876–894
126–147	39	22	61	854–875
148–171	40	24	60	830–853
172–198	41	27	59	803–829
199–227	42	29	58	774–802
228–258	43	31	57	743–773
259–291	44	33	56	710–742
292–326	45	35	55	675–709
327–363	46	37	54	638–674
364–401	47	38	53	600–637
402–440	48	39	52	561–599
441–480	49	40	51	521–560
481–520	50	40		

(ii) The concept of degrees of freedom arises in the application of the t-test. The number of degrees of freedom is the number of independent variables from which a statistic is computed. In computing the mean of an n-fold sample there are n degrees of freedom; in computing the variance of the sample we have to sum $\Sigma\,(x_r - m)^2$ where $\Sigma\,(x_r - m) = 0$, i.e. the n variables $(x_r - m)$ are constrained by the relation that their sum is zero, and one degree of freedom is lost. Hence the 'unbiassed' estimate of the variance of a sampled population is $\Sigma\,(x_r - m)^2/(n-1)$ if m is estimated from the sample, or $\Sigma\,(x_r - \mu)^2/n$ if μ is the known mean of the sampled population.

The concept becomes clearer if very small samples are considered. A sample of one gives an estimate of the mean but no information about the variance unless the mean is already known: a sample of at least two is needed to give an estimate of the variance if the mean is unknown.

(iii) For sampling tests with small samples an artificial Normal distribution is useful. It can consist merely of a Normal frequency distribution to be used with a table of random numbers, or of a set of counters each marked with a number which can be sampled as required (Table 8).

4. Bivariate Distributions (*Section I; YK, Ch.* 9; *W, Ch. IV*)

4(*a*) *Scatter Diagrams*

(i) The problem of representing bivariate distributions by methods analogous to those used for univariate distributions needs brief discussion. The 'solid histogram' representing a grouped bivariate distribution and its 2-dimensional representation by a 'dot-diagram' should be shown.

(ii) Meanings for the terms *unassociated variables, variables with linear association, variables associated non-linearly* are established by illustrating them by appropriate scatter diagrams and examples.

4(*b*) *Linear Regression*

(i) The lines of regression are defined as the loci of the means

of the two sets of arrays (i.e. the rows and columns) of a grouped frequency distribution. In small samples the means are rarely exactly linear and some method of estimating the lines is required. A graphical method is described in Section I.

(ii) The direct calculation of the equations of regression lines involves the computation of the *covariance*, or first product-moment, of the distribution. It is given by

$$\Sigma\,(x_r - m_x)(y_r - m_y)/(n - 1)$$

and some of its algebraic properties should be established before its computation is explained. One property that will be needed is that

$$\Sigma\, x_r y_r = \Sigma\, X_r Y_r - n\,ab$$

where x_r, y_r, are the deviations from the true means; X_r, Y_r, are the deviations from arbitrary means, and a, b, are the deviations of the true means from the arbitrary means.

(iii) The deviation of the formulae of the regression coefficients using the Principle of Least Squares is given in W, p. 70, and avoiding calculus methods in YK, p. 216.

(iv) Before computing the coefficients it is important to transform the data to the simplest possible form for computation (and to transform the equations back again after the computation).

(v) The following variations of the computation occur:

(*a*) Direct computation of both lines from ungrouped data (YK, p. 224).

(*b*) Computation of both lines from grouped data, (i) making individual entries of the deviation products for each bivariate group (YK, p. 226), (ii) using the 'diagonal' method based on the identities

$$\Sigma\,(x \pm y)^2 = \Sigma\,(x^2) + \Sigma\,(y^2) \pm 2\,\Sigma\,(xy)$$

by which the products are computed using tables of squares (YK, p. 229).

(*c*) Computation of a linear trend line for time-series.

(vi) Though significance tests of regression coefficients are not required in a school course it should be pointed out that they are subject to sampling errors.

4(*c*) *The Product-Moment Correlation Coefficient*

(i) The product-moment coefficient of correlation, r, is a measure of the *linear* association of two variables. A low value of $| r |$ does not necessarily imply a lack of association; a high value of $| r |$ does not imply a causal relationship between the two variables.

(ii) The coefficient is subject to sampling errors which in small samples can be very large, e.g. the minimum values of $| r |$ that are significant at the 5% level for 10, 20, 50, 100 pairs of values drawn from a Normal bivariate population are 0·63, 0·44, 0·27 and 0·20 respectively.

(iii) Some properties of r should be established, e.g., (*a*) that it is unaffected by linear transformations of the variables, (*b*) that it is the geometric mean of the regression coefficients.

4(*d*) *The Rank Correlation Coefficient*

(i) The rank correlation coefficient is obtained as the product-moment correlation coefficient of the paired ranks of the data instead of the crude data. The derivation of the simplified formula and the procedure to be adopted when tied ranks occur are given in YK, pp. 261, 264.

(ii) The probability distribution of the coefficient for small samples (no ties) provides interesting algebraic exercises.

(iii) Algebraic proof that the coefficient ranges from $+ 1$ to $- 1$ forms a simple exercise.

5. Simple Contingency Tables (*YK*, *Ch.* 1, 2)

(i) The syllabus item 'simple contingency tables' occurs under the heading 'Two-variate distributions' and it is therefore assumed that the theory of 2×2 contingency tables will suffice.

(ii) A first need is to establish the notation (following YK):

(*a*) A and B designate attributes; α and β the absence of these attributes.

(*b*) (A) denotes the number of items with the attribute A; (AB) the number with both attributes, and so on.

(iii) On the hypothesis of independence, expected frequencies are given by the entries in the table:

Attribute	A		Totals
B	$(A)(B)/N$	$(\alpha)(B)/N$	(B)
β	$(A)(\beta)/N$	$(\alpha)(\beta)/N$	(β)
Totals . . .	(A)	(α)	N

(iv) Allowance for sampling fluctuations must be made in any discussion of results, but as no criteria for testing are required the expected and observed frequencies must differ markedly before any association can be presumed and the samples must be large.

(v) A simple measure of association is helpful in discussion. That provided by:

$$Q = \frac{(AB)(\alpha\beta) - (\alpha B)(A\beta)}{(AB)(\alpha\beta) + (\alpha B)(A\beta)}$$

is suitable. It is zero if the attributes are independent and ranges from $+1$ (complete positive association) to -1 (complete negative association).

(vi) An example: In an examination of 40 boys 22 passed in both Mathematics and Physics, 7 failed in both subjects and 5 failed in Mathematics only. Complete the table, compare these frequencies with those expected on the hypothesis of complete independence, and comment. The complete table is:

	Mathematics		
Physics	Pass	Fail	Totals
Pass	22 [19]	5 [8]	27
Fail	6 [9]	7 [4]	13
Totals	28	12	40

The expected frequencies are $27 \times 28/40$, $12 \times 27/40$, $28 \times 13/40$ and $12 \times 13/40$, which to the nearest whole numbers give the numbers in square brackets. The value of Q is $(22 \times 7 - 5 \times 6)/(22 \times 7 + 5 \times 6) = 0{\cdot}67$, which suggests a fairly strong positive association between passes in Mathematics and passes in Physics.

This example, with admittedly a small sample, illustrates the difficulty of deciding without some criterion or some experience how big a deviation from the expected value must be before it can be stated with confidence that it is unlikely to be a random fluctuation.

INDEX

Continued from back cover

SOME OPINIONS ON

"AN ... ETHOD"

...report on "The ... of modern statis- ... W. F. L. Dick

...course adequate ... in Education; ... foundation ... much

...ever, ... many ... statis- ... with

...s has ... f the ... omed ... g the

...pula- ... sition ... tions. ... n and ... n the ... s and

... clear ... erage ... rence ... book

...ion, is ... more ... uency ... n and ... many ... exer- ... ests of significance, regression and correlation. The teacher who is introducing statistics in his classes will find all he needs for himself or for his sixth form pupils in this book. The beginner at research, whether in the field of education or elsewhere, cannot afford to be without it.

A School Textbook of Statistics, published in 1951

AN INTRODUCTION TO STATISTICAL METHOD

By

B. C. Brookes, M.A.

Lately Senior Mathematics Master, Bedford Modern School.
Lecturer in the Presentation of Technical Information,
University College, London.
Member of the Teaching Committee of the
Royal Statistical Society

18

and

W. F. L. Dick, M.A.

Technical Officer, Imperial Chemical Industries

This elementary introduction has been written to cover all school requirements. It is also useful for anyone who needs a simple and readable introduction to the subject. *A selection of opinions is given here on the inside cover.*

Unlike many other "introductions" this book really begins at the beginning. It ranges from Descriptive Statistics to the Design of Experiments. Very little previous mathematical knowledge is required (G.C.E. Ordinary Level—though some elementary calculus methods are used).

A special feature is the collection of over 300 exercises and examples, with answers and hints for solution. Many of these are not artificially prepared, but are actual examples taken from the many fields of activity where statistics are used today. They include practical experiments which can be carried out by groups of students.

The authors have a wide experience, both of teaching statistical methods at an elementary level and of applying them in practice.

Available in Two Editions:

Parts I and II together: 21s. net. *Part I alone:* 9s. 6d.

For Opinions, see inside cover

WM. HEINEMANN LTD. 99 GT. RUSSELL ST., W.C.1